

+ (51) 298

⑫

EUROPEAN PATENT APPLICATION

⑮ Application number: 84304801.8

⑯ Date of filing: 13.07.84

⑰ Int. Cl.⁴: **C 12 N 15/00**
C 12 P 21/00
//C12R1/19, C12R1/05,
C12R1/465, C12R1/07, C12R1/06,
C12R1/865

⑳ Priority: 15.07.83 GB 8319265
13.03.84 GB 8406569

㉑ Date of publication of application:
19.06.85 Bulletin 85/25

㉒ Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

㉓ Applicant: **The University of Warwick**
Coventry West Midlands, CV4 7AL(GB)

㉔ Inventor: **Lord, John Michael**
93 Campion Court
Leamington Spa Warwickshire(GB)

㉕ Inventor: **Roberts, Lynne Margaret**
48 Braemar Road Liffington
Leamington Spa Warwickshire(GB)

㉖ Inventor: **Lamb, Francis Ian**
76 Henry Street
Kenilworth Warwickshire CV8 2HJ(GB)

㉗ Representative: **Bannerman, David Gardner**
Withers & Rogers 4 Dyer's Buildings Holborn
London, EC1N 2JT(GB)

㉘ DNA sequence coding for a plant toxin of the ricin type, or a portion thereof.

㉙ DNA comprising a nucleotide sequence coding for at least a portion of the precursor polypeptide of ricin or a similar plant toxin. Recombinant DNA molecules containing such DNA as an insert, and genetically-modified host microorganisms containing such recombinant DNA molecules.

DGB/JEA

-1-

DNA

This invention relates to DNA comprising a nucleotide sequence coding for at least a portion of a polypeptide which is a plant toxin of the ricin type, as herein-after defined. It also relates to recombinant DNA

5 molecules comprising a DNA sequence which codes for a polypeptide which is or is closely related to a natural plant toxin of the ricin type. Ricin, and also other plant toxins such as abrin, modeccin and viscumin, consist of two polypeptide chains (known as the A and B

10 chains) linked by a disulphide bridge, one chain (the A chain) being primarily responsible for the cytotoxic and the other chain (the B chain) having sites enabling the molecule to bind to cell surfaces. Ricin is produced in the plant Ricinis communis (also known as the castor

15 bean plant) via a precursor protein known as "preproricin".

Preproricin comprises a single polypeptide chain which includes a leader sequence. The leader sequence is subsequently removed in the organism to give proricin

20 which is then cleaved to eliminate a linker region and joined by a disulphide bond to form the mature protein.

The toxicity of ricin-type toxins operates in three phases: (1) binding to the cell surface via the B chain; (2) penetration of at least the A chain into the cytosol, and (3) inhibition of protein synthesis through the A chain attacking the 60S subunits of the ribosomes. Thus, separated A and B chains are essentially non-toxic, the inherently toxic A chain lacking the ability to bind to cell surfaces in the absence of the B chain.

It is also known that in ricin-type toxins the B chain binds to cell surfaces by virtue of galactose recognition sites, which react with glycoproteins or glycolipids exposed at the cell surface.

It has already been suggested that the toxicity of the ricin A chain might be exploited in anti-tumour therapy, by replacing the indiscriminately-binding B chain with a different carrier component having the ability to bind only to tumour cells. Thus, various immunotoxins have already been prepared, consisting of a conjugate of whole ricin or a separated natural ricin A chain and a tumour-specific monoclonal antibody. Although these known conjugates are of considerable potential in themselves, there is scope for improvement.

One problem with the known conjugates arises from a structural feature of the A chain from natural ricin. It is known that the natural ricin A chain becomes N-glycosylated during its synthesis, by enzymes present

in Ricinus cells, and it is thought that the resulting sugar moieties are capable of non-specific interactions with cell surfaces. Thus, it appears that the known A chain conjugates are capable of a certain amount of binding with non target cells, even in the absence of the natural B chain, thus increasing the toxicity of such immunotoxins towards non target cells.

Another problem with the known ricin A chain conjugates stems from the fact that the B chain seems to have an important secondary function in that it somehow assists in the intoxication process, apart from its primary function in binding the ricin molecule to the cell surface. This secondary function is lost if the B chain is replaced by a different carrier component such as a monoclonal antibody.

If it were possible to prevent interactions between the cell surface via the A chain sugar moieties, whilst preserving the secondary toxicity-increasing function of the B chain, the toxicity of a whole ricin antibody conjugate towards normal cells could be reduced, and towards target cells could be increased, thus improving the therapeutic index of the immunotoxin. It is also known that the natural ricin B chain is N-glycosylated and the B chain sugar moieties may also contribute to non specific interactions. Also, the sugar moieties in both chains enable the ricin molecule to be sequestered by reticuloendothelial cells in the liver, and so would lead to the

rapid excretion from the system of a drug based on a part of the whole of the ricin molecule in which such sugar moieties were still present.

Attempts to remove all the sugar moieties from natural ricin by chemical or enzymatic methods have so far failed. Nevertheless the major obstacle confronting the use of known whole ricin-antibody conjugates is the presence of two galactose binding sites in the ricin B chain. These B chain galactose binding sites are primarily responsible for the non-specific cellular interactions of current whole ricin-antibody conjugates, particularly when used in vivo. Their presence in the natural toxin clearly eliminates or reduces the targeting specificity conferred by the antibody.

An improved immunotoxin based on ricin or another plant toxin of the ricin type, not suffering from these problems, could consist of a whole toxin molecule modified so that it is not N-glycosylated, and so that the B chain has no galactose recognition sites, but retains its secondary intoxication-promoting properties, coupled to a carrier moiety which delivers the toxin to the target cells. This could be a tumour-specific or cell/tissue specific vehicle such as a suitable monoclonal antibody.

Our research which has so far been concentrated on ricin itself, has indicated that the assembly of ricin (and the related agglutinin which consists of

two ricin-like molecules with slightly modified A and B chains) do not involve the separate synthesis of the A and B chains as the products of distinct mRNA's,

but rather the initial formation of a single polypeptide precursor containing both the A chain and B chain sequences. This is thought to apply in the case of other toxins of the same type.

This invention is based on the idea of preparing a genetically-engineered microorganism capable of expressing a molecule of a toxin of the ricin type, as defined above,

or alternatively part of such a toxin molecule, or a precursor of such a molecule (which could be converted to the toxin molecule itself) which toxin molecule could be modified as suggested above and could be used to construct an effective toxin conjugate by combining it with a tumour-specific or cell/tissue specific monoclonal antibody or other carrier moiety, such as a hormone or lectin.

The fact that ricin is formed via a precursor polypeptide will enable a cell system to be constructed by known techniques which expresses a ricin precursor. The ricin precursor product could then be chemically or enzymatically converted to the desired modified ricin. An analogous technique could be used in the case of other ricin-type toxins as herein defined. An alternative technique would be to divide from the DNA sequence that codes

for the precursor two sequences which code separately
for the A and B chains, to insert these separated
sequences into different cloning vehicles and to insert
the resulting recombinant DNA molecules into separate
5 host microorganisms. One such host would then express
the A chain polypeptide sequence and the other the B
chain polypeptide sequence. These sequences could
then be combined to form the desired modified ricin
molecule. This technique could obviously also be used
10 for any ricin-type toxins in which the A and B chains
are encoded by distinct mRNA gene pools. This latter
approach would be preferred on safety grounds, in that
separate and therefore non-toxic A and B chains would
be expressed.

15 According to one aspect of the invention we provide
a biologically pure and homogeneous sample of DNA com-
prising a nucleotide sequence coding for at least a
portion of a precursor of a ricin-type toxin polypeptide,
or mutants thereof.

20 Said portion preferably comprises the A chain
or the B chain of the mature protein.

More specifically, we provide a sample of
DNA including at least a substantial portion of any
of the following DNA sequences, which sample is
25 biologically-pure:

ATG TAT GCA GIG GCA ACA TGG CIT

TGT TTT GGA TCC ACC TCA GGG TGG TCT TTC ACA TTA GAG GAT AAC AAC ATA

TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGT GTC ACT GIG

CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TTA ACA ACT GGA

GCT GAT GIG AGA CAT GAT ATA CCA GIG TIG CCA AAC AGA GTT GGT TIG CCT

⁵ATA AAC CAA CGG TTT ATT TTA GTT GAA CTC TCA AAT CAT GCA GAG CTI TCT

GTT ACA TTA GGC CTG GAT GTC ACC AAT GCA TAT GIG GTC GGC TAC CGT GTT

GGA AAT AGC GCA TAT TTC TTT CAT CCT GAC AAT CAG GAA GAT GCA GAA GCA

ATC ACT CAT CTI TTC ACT GAT GTT CAA AAT CGA TAT ACA TTC GCC TTT GGT

Page 7a follows...

GGT AAT TAT GAT AGA CTT GAA CAA CTI GCI GGT AAT CTG AGA GAA AAT ATC

GAG TTG GGA AAT GGT CCA CTA GAG GAG GCI ATC TCA GCG CTI TAT TAT TAC

AGT ACT GGT GGC ACT CAG CTT CCA ACT CTG GCT CGT TCC TTT ATA ATT TGC

ATC CAA ATG ATT TCA GAA GCA GCA AGA TTC CAA TAT ATT GAG GGA GAA ATG

5 CGC ACG AGA ATT AGG TAC AAC CGG AGA TCT GCA CCA GAT CCI AGC GTA ATT

ACA CTI GAG AAT AGI TGG GGG AGA CTI TCC ACT GCA ATT CAA GAG TCT AAC

CA4 GGA GGC TTI GCI AGT CCA ATT CAA CTG CAA AGA CGT AAT GGI TCC AAA

TTT AGT GTG TAC GAT GTG AGT ATA TTA ATC CCI ATC ATA GCI CTC ATG GTG

TAT AGA TGC GCA CCT CCA CCA TCG TCA CAG TTI TCT TTG CTI ATA AGG CCA

Page 7b follows...

GTG GTA CCA AAT TTT AAT GCT GAT GTT TGT ATG GAT CCT GAG CCC ATA GTC

CGT ATC GTA GGT CGA AAT GGT CTA TGT GTT GAT GTT AGG GAT GGA AGA TTC

CAC AAC GGA AAC GCA ATA CAG TTG TGG CCA TGC AAG TCT AAT ACA GAT GCA

AAT CAG CTT TGG ACT TTG AAA AGA GAC AAT ACT ATT CGA TCT AAT GGA AAG

5. TGT TTA ACT ACT TAC GGG TAC AGT CCG GGA GTC TAT GTG ATG ATC TAT GAT

TGT AAT ACT GCT GCA ACT GAT GCC ACC CGC TGG CAA ATA TGG GAT

AAT GGA ACC ATC ATA AAT CCC AGA TCT AGT CTA GTT TTA GCA GCG ACA TCA

GGG AAC AGT GGT ACC ACA CTT ACG GTG CAA ACC AAC ATT TAT GCC GTT AGT

CAA GGT TGG CTT CCT ACT AAT AAT ACA CAA CCI TTT GTT ACA ACC ATT GTT

Page 7c follows...

0145111

GGG CTA TAT GGI CTG TGC TTG CAA GCA AAT AGT GGA CAA GTA TGG ATA GAG

GAC TGT AGC AGT GAA AAG GCT GAA CAA CAG TGG GCT CTT TAT GCA GAT GGT

TCA ATA CGT CCT CAG CAA AAC CGA GAT AAT TGC CTT ACA AGT GAT TCT AAT

ATA CGG GAA ACA GTT GTT AAG ATC CTC TCT TGT GGC CCT GCA TCC TCT GGC

5. CAA CGA TGG ATG TTC AAG AAT GAT GGA ACC ATT TTA AAT TIG TAT AGT GGA

TTG GTG TTA GAT GTG AGG CGA TCG GAT CCG AGC CTT AAA CAA ATC ATT CTT

TAC CCT CTC CAT GGI GAC CCA AAC CAA ATA TGG TTA CCA TTA TTT

OT

ATG TAT GCA GTG GCA ACA TGG CTT

TGT TTT GGA TCC ACC TCA GGG TGG TCT TTC ACA TTA GAG GAT AAC AAC ATA

TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGT GGC ACT GTG

CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TTA ACA ACT GGA

GCT GAT GTG AGA CAT GAT ATA CCA GTG TTG CCA AAC AGA GTT GGT TTG CCT

5. ATA AAC CAA CGG TTT ATT TTA GTT GAA CTC TCA AAT CAT GCA GAG CTT TCT

GTT ACA TTA GCC CTG GAT GTC ACC AAT GCA TAT GTG GTC GGC TAC CGT GCT

GGA AAT AGC GCA TAT TTC TTT CAT CCT GAC AAT CAG GAA GAT GCA GAA GCA

ATC ACT CAT CTT TTC ACT GAT GTT CAA AAT CGA TAT ACA TTC GCC TTT GGT

Page 8a follows...

GGT AAT TAT GAT AGA CTT GAA CAA CTT GCT GGT AAT CTG AGA GAA AAT ATC

GAG TTG GGA AAT GGT CCA CTA GAG GAG GCT ATC TCA GCG CTT TAT TAT TAC

AGT ACT GGT GGC ACT CAG CTT CCA ACT CTG GCT CGT TCC TTT ATA ATT TGC

ATC CAA ATG ATT TCA GAA GCA GCA AGA TTC CAA TAT ATT GAG GGA GAA ATG

5. CGC ACG AGA ATT AGG TAC AAC CGG AGA TCT GCA CCA GAT CCI AGC GTA ATT

ACA CTT GAG AAT AGT TGG GGG AGA CTT TCC ACT GCA ATT CAA GAG TCT AAC

CAA GGA GCC TTT GCT AGT CCA ATT CAA CTG CAA AGA CGT AAT GGT TCC AAA

TTT AGT GTG TAC GAT GTG AGT ATA TTA ATC CCT ATC ATA GCT CTC ATG GTG

TAT AGA TGC GCA CCT CCA CCA TCG TCA CAG TTT

or

Page 8b follows

GCT GAT GTT TGT ATG GAT CCT GAG CCC ATA GTG

CGT ATC GTA GGT CGA AAT GGT CTA TGT GTT GAT GTT AGG GAT GGA AGA TTC

CAC AAC GGA AAC GCA ATA CAG TTG TGG CCA TGC AAG TCT AAT ACA GAT GCA

AAT CAG CTC TGG ACT TTG AAA AGA GAC AAT ACT ATT CGA TCT AAT GGA AAG

5. TGT TTA ACT ACT TAC GGG TAC AGT CCG GGA GTC TAT GTG ATG ATC TAT GAT

TGC AAT ACT GCT GCA ACT GAT GCC ACC CGC TGG CAA ATA TGG GAT

AAT GGA ACC ATC ATA AAT CCC AGA TCT AGT CTA GTT TTA GCA GCG ACA TCA

GGG AAC AGT GGT ACC ACA CTT ACG GTG CAA ACC AAC ATT TAT GCC GTT AGT

CAA GGT TGG CTT CCT ACT AAT AAT ACA CAA CCT TTT GTT ACA ACC ATT GTT

GGG CTA TAT GGT CTG TGC TTG CAA GCA AAT AGT GGA CAA GAA TGG ATA GAG

GAC TGT AGC AGT GAA AAG GCT GAA CAA CAG TGG GCT CTT TAT GCA GAT GGT

TCA ATA CGI CCI CAG CAA AAC CGA GAT AAT TGC CTT ACA AGT GAT TCT AAT

ATA CGG GAA ACA GTT GTT AAG ATC CTC TCT TGT GGC CCT GCA TCC TCT GGC

5•CAA CGA TGG ATG TTC AAG AAT GAT GGA ACC ATT TTA AAT TTG TAT AGT GGA

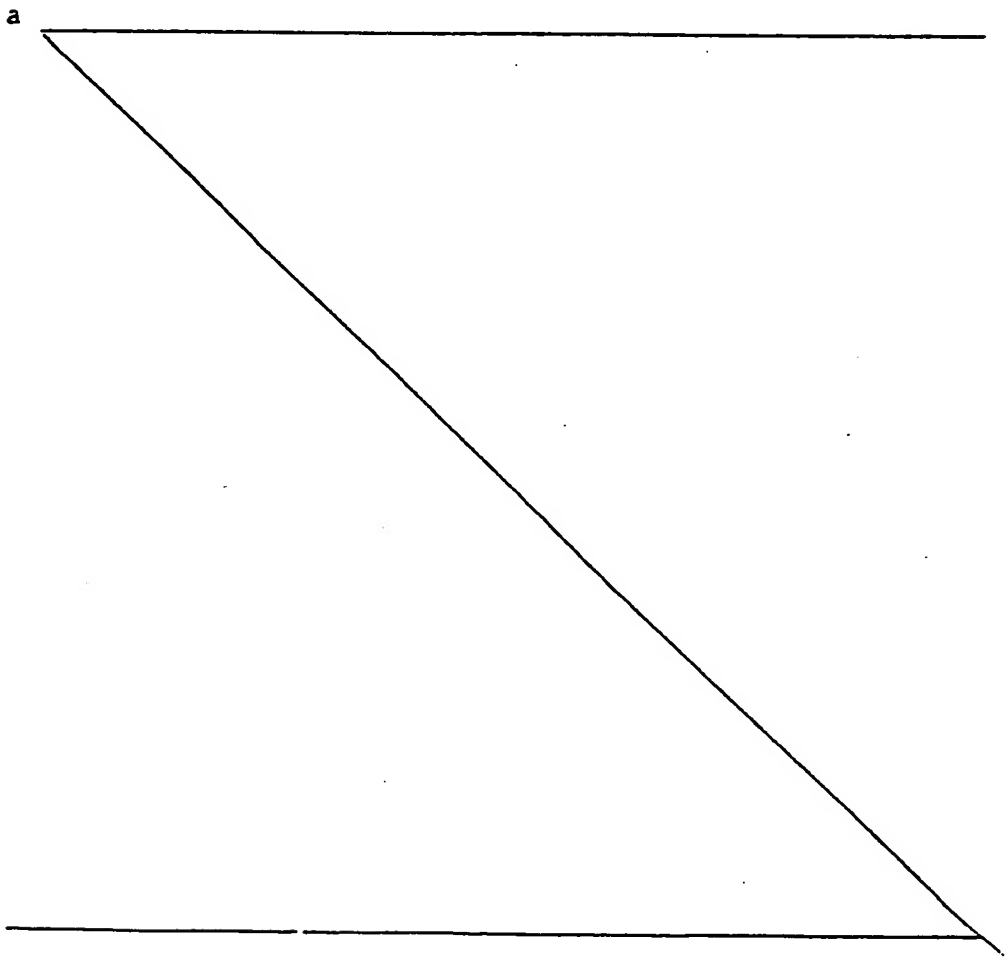
TTG GTG TTA GAT GTG AGG CGA TCG GAT CCG AGC CTT AAA CAA ATC ATT CTT

TAC CCI CTC CAT GGT GAC CCA AAC CAA ATA TGG TTA CCA TTA TTT

or at least a portion of a nucleotid sequence which is equivalent thereto by virtue of degeneracy of the genetic code.

According to another aspect of this invention
5 we provide a recombinant DNA molecule comprising a DNA sequence which codes for a polypeptide sequence present in a plant toxin of the ricin type, as defined herein.

More specifically, we provide a recombinant DNA
10 molecule containing a DNA sequence which codes for



A and B chain precursor polypeptide of a plant toxin of the ricin typ .

Alternatively, we provide a recombinant DNA molecule containing a DNA sequence coding for at least
5 a portion of either the A chain or the B chain of a plant toxin of the ricin type.

According to a further aspect of the present invention we provide a genetically modified host micro-organism which contains a recombinant DNA molecule as
10 hereinbefore defined.

In the recombinant DNA molecules according to the present invention, the nucleotide sequence coding for the B chain may be modified to eliminate or inactivate the galactose binding sites, and the precursor polypeptide
15 and hence in the mature protein, to eliminate or inactivate the galactose binding sites, and the sequences encoding ther signals for N-glycosylation may also be modified to render them ineffective or to eliminate them. Examples of techniques which may prove useful are deletion or
20 oligonucleotide mediated mutagenesis.

The host organism may be a plant cell or an animal cell or preferably a microorganism.

The microorganism may be a prokaryote or a eukaryote. As examples of prokaryotes may be mentioned
25 Gram-negative bacteria, e.g. E. coli., Methylophilus methylophilus and Alcaligenes eutrophus; and Gram-positiv

-11-

bact ria, e.g. Streptomyces, Bacillus subtilis and Arthrobacter. As examples f eukaryot s may be mentioned yeasts, for example Saccharomyces cerevisiae.

The recombinant DNA molecules may comprise a cloning
5 vector such as a plasmid or phage vector into which has been inserted the DNA sequence coding for at least a portion of a precursor polypeptide, or at least a portion of either the A chain or the B chain, of a ricin-type plant toxin.

10 The cloning vector is preferably a plasmid although we do not exclude the possibility that it may be a phage vector. The plasmid may be a naturally-occurring plasmid or preferably a composite derived from fragments of other plasmids. Where a composite plasmid is
15 employed it preferably contains promoter sequences which improve expression of the ricin gene.

Examples of suitable plasmids which may be used as cloning vehicles are inter alia for Gram-negative bacteria: pBR322, pAT153, pUC8, pGSS15 and pMB9; for
20 Gram-positive bacteria: pVC6; and for S.cerevisiae: pMA91, pMA230, YRp7, pLC544 and YEp6. The vector will be selected to be suitable for the particular host envisaged.

We also provide a method of obtaining a recombinant DNA molecule which comprises preparing a double-stranded DNA
25 sequence which codes for a polypeptide sequence present in a plant toxin of the ricin type and inserting said double-stranded DNA sequence into a cloning vector.

More specifically such a method may comprise isolating the mRNA which codes for the ricin A and B chain precursor polypeptide, synthesising single-stranded cDNA from said mRNA using reverse transcriptase and a suitable primer, assembling the second DNA strand onto the template formed by said first strand by means of DNA polymerase followed by S1 nuclease, and inserting the resulting double-stranded cDNA into a cloning vector.

Alternatively, the cDNA assembled from the mRNA may be cut into separate portions which code respectively for separate portions of the ricin molecule, for example for the A and B chains, which portions are then inserted into separate cloning vectors.

As stated above the cloning vector is preferably a plasmid such as pBR322, pAT153, or pUC8, and this may be cut open by means of the restriction endonuclease Pst I, and may be tailed with oligo (dG), and annealed with the double-stranded cDNA which has been tailed with oligo (dC).

We also provide a method of producing a modified transformed host by introducing into a suitable host microorganism a recombinant DNA molecule according to this invention.

The microorganism used as the host for cloning is preferably a Gram-negative bacterium and more preferably E. coli.

After cloning, the DNA sequence coding for the ricin precursor (or the precursor of another ricin-type

-13-

toxin which is formed from a precursor) may be removed from the host cloning vector. It may then be divided into two portions which code for separate regions of the toxin molecule, for example the A and B chains, these portions introduced into separate second cloning vectors and new hosts modified with each of the resulting new recombinant DNA molecules. Alternatively it may be introduced whole into a second cloning vector. The second cloning vectors comprise suitable promoter sequences and the position and direction of insertion of the whole coding sequence or portion thereof into the second cloning vector are such that on introducing the new recombinant DNA molecules into suitable host microorganisms, e.g. E. coli or S.cerevisiae, expression of the desired gene sequence is obtained.

The preparation of a transformed host containing a DNA sequence which codes for the ricin A and B chain precursor polypeptide will now be described by way of example, first in general terms and then in detail. This process is summarised in the accompanying diagram.

Firstly, the mRNA encoding this precursor was enriched in known manner by sucrose density gradient centrifugation. The corresponding cDNA was assembled in single-strand form onto this mRNA in known manner using the enzyme reverse transcriptase, a growing point having first been provided on the mRNA using oligo (dT) as primer which binds onto the polyadenylated 3'-terminus

f th mRNA. The immediate product of this reaction is a DNA-RNA hybrid. The RNA strand is removed by hydrolysis, leaving the single-strand DNA intact. This is converted to the double stranded form using the enzyme DNA polymerase in the presence of free nucleotides, which results in a hairpin-shaped molecule, the curved end of which is then removed by the single-strand specific nuclease S1. The resulting double strand cDNA is then tailed with oligo (dC)/, ^{using terminal transferase} size fractionated to remove small molecules, or vice versa and annealed with the pBR322 or pAT153 vector which has been cut open with Pst I and tailed with oligo (dG)/, ^{using terminal transferase} the cytosine tails on the DNA base pairing with the guanine tails on the vector.

15 The resulting chimaeric plasmids containing the DNA segment coding for the ricin precursor polypeptide were then used to transform E-coli DH1 cells, and the presence of the chimaeric plasmid was ensured by selecting cells displaying tetracycline resistance and ampicillin sensitivity. Over 1600 Tet^r, Amp^s clones were obtained. Colonies derived from each clone were transferred to nitrocellulose filters and clones containing the desired DNA sequence identified using a 32P-end labelled 20 mer oligonucleotide probe, having the DNA sequence ACCTACAA^ATT^CTT^ACT^ACC^G which hybridises to DNA containing the complementary sequence TGGATGTT^TAA^GAA^TGA^TGG^C. As the ricin precursor

polypeptide have been found to contain the amino sequence - Trp-Met-Phe-Lys-Asn-Asp-Gly- the DNA sequence responsible for this is known from the genetic code to be the latter mentioned above.

Using appropriate hybridisation and wash conditions, e.g. as described by Singer-Sam et al in (1983) Proc. Natl.Acad.Sci.(U.S.A.), Vol.80 pp 802-806, 80 clones were selected as positively containing the desired DNA sequence, and of these, the eight largest in the plasmid pBR322 have been initially chosen for further characterisation. Their relationship to the castor bean lectin precursor polypeptides has been confirmed using the hybrid release translation assay. Of the eight clones mentioned above, four, respectively with 1614, 1950, 1059 and 1020 base pairs, have been selected for sequencing.

In detail, the transformed host was prepared as follows:

A. cDNA synthesis

1. mRNA extraction and fractionation

100-200 g of ripening Ricinus seeds were frozen and ground to a powder in liquid nitrogen, and homogenised in a Waring blender for 1 - 2 minutes in 50 mM tris-HCl pH 9, 150 mM NaCl, 5 mM EDTA and 5% SDS. The homogenate was extracted with an equal volume of phenol:chloroform (1:1) and the phases were separated by centrifugation. The organic phase and residue

-16-

were reextracted with 0.5 volume of 20 mM tris-HCl
pH 9.0, 2 mM EDTA and the resultant aqueous phase
was combined with the original one. The total
aqueous phase was reextracted repeatedly with equal volumes
5 of phenol:chloroform until no material was present
at the interface. RNA was precipitated by the
addition of 2 volumes of cold ethanol after making
the solution 200 mM in NaCl.

After overnight precipitation at -20°C the RNA
10 was centrifuged at 10,000 rpm for 30 minutes in
an MSE 18 or MSE 21 centrifuge; the pellet was
then washed repeatedly in 3 M NaAc pH 5.5 until
no polysaccharide was detectable in the supernatant
by ethanol precipitation. The final pellet was
15 dissolved in 300 mM NaCl and precipitated as above.

mRNA molecules bearing poly(A) tails were extracted
by affinity chromatography on oligo(dT)-cellulose:
after hybridisation at room temperature for 30 min.
in 400 mM NaCl, 20 mM tris-HCl pH 7.6, 0.2% SDS,
20 the beads were pelleted and washed three times in
the above buffer and two times in 200 mM NaCl,
20 mM tris-HCl pH 7.6, 0.1% SDS. The slurry was
poured into a column and washed further with the
last buffer until the A_{260} of the eluate reached
25 the background level. Poly(A)-containing RNA was
then eluted with 20 mM tris-HCl pH 7.6 at 50°C .
The eluate was monitored with an ISCO continuous
flow UV cell. Poly(A)-containing RNA was precipitated
overnight from 200 mM NaCl by the addition of 2
30 volumes of cold ethanol at -20°C and was then washed

three times with 70% ethanol, and redissolved in 10 mM tris-HCl pH 7.0 to approximately 1 $\mu\text{g}/\mu\text{l}$.

The mRNA was heated for 2 min. at 65°C and quench cooled. Approximately 400 μg of poly(A)⁺ RNA was
5 layered on top of a 10 - 30% ribonuclease-free sucrose (Sigma) density gradient in 100 mM tris-HCl pH 7.5, 0.5% SDS, 1 mM EDTA, and centrifuged in a Beckman 15-65B centrifuge using an SW27 rotor at 25,000 rpm at 17°C for 14 hours. 400 μl fractions were
10 collected with an ISCO density gradient fractionator using the continuous flow UV cell.

Each fraction was made 200 mM in NaCl and precipitated with 2 volumes of cold ethanol by freeze-thawing in liquid nitrogen three times, and recovered by
15 centrifugation in an Eppendorf microcentrifuge for 30 minutes at 4°C, washed once with 70% ethanol, and redissolved in 10 μl of 10 mM tris-HCl pH 7.0. An aliquot from each fraction (1 μl) was translated in a reticulocyte lysate cell free system and the
20 lectin precursor immunoprecipitated to identify the fraction enriched for lectin mRNA.

2. First strand synthesis

Fractionated poly(A)⁺ RNA was reverse transcribed at 0.5 $\mu\text{g}/\mu\text{l}$ in the presence of 50 mM tris-HCl pH
25 8.3, 10 mM MgCl_2 , 100 mM KCl, 1 mM of dATP, dTTP and dGTP, 250 μM dCTP, 0.06 $\mu\text{g}/\mu\text{l}$ olig (dT)₁₂₋₁₈, 10 mM DTT and 0.4 units/ μl of reverse transcriptase

from avian myeloblastoid virus. (3H)dCTP or
Q-(³²P)dCTP were included in the reaction as appropriate.

The reaction mixture was incubated at 42°C for 45
minutes, at which point an equal volume of 5 mM
5 tris-HCl pH 8.3, 5 mM DTT, 250 μM dCTP was added
along with the same amount of enzyme as previously.
The reaction was incubated for a further 45 minutes
at 45°C and terminated by freezing. Aliquots were
analysed on 1% denaturing agarose gels along with
10 the products of the second strand and S₁ nuclease
reactions.

3. Second strand synthesis

mRNA-cDNA hybrids were denatured by boiling the
first strand reaction for 3 minutes and cooling
15 rapidly. After pelleting insoluble material in
the Eppendorf microfuge for 2 minutes the supernatant
was transferred to a fresh chilled tube. For the
standard reaction, reagents were added as follows,
ignoring elements already present: dATP, dGTP and
20 dTTP to 100 μM, Hepes-KOH pH 6.9 to 105 mM, KCl
to 92 mM, dCTP, labelled as appropriate, to 80 μM,
and 0.1 units/μl of DNA polymerase. The reaction
was allowed to proceed at 20°C for 6 hours, at which
time cDNA was removed from the mixture by gel filtration
25 on 1 ml columns of Bio-Gel P60 in 10 mM tris-HCl
pH 7.6, 20 mM NaCl, 1 mM EDTA. Fractions were
monitored by Cerenkov or liquid scintillation counting,

and peak excluded fractions were pooled and precipitated from 0.3 M NaAc pH 6 by the addition of 2 volumes of cold ethanol. Precipitates were recovered by centrifugation in the Eppendorf micro-
5 centrifuge for 30 minutes in the cold, and dissolved in water to about 2.5 $\mu\text{g}/\mu\text{l}$ of RNA - equivalent material.

4. S₁ nuclease digestion

Single-stranded regions of double-stranded cDNA
10 were digested with S₁ nuclease from Aspergillus oryzae, in the presence of 300 mM NaCl, 30 mM NaAc pH 4.5, 3 mM ZnCl₂. The reaction was incubated for 15 minutes at 37°C and then for 15 minutes at 15°C, and was terminated by the addition of tris-
15 HCl pH 7.6 to 130 mM and EDTA to 10 mM; it was then extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated from 300 mM NaAc pH 6 with 2 volumes of cold ethanol. The precipitate was dissolved in 10 mM tris-HCl
20 pH 8, 0.1 mM EDTA to 0.25 $\mu\text{g}/\mu\text{l}$ RNA equivalent.

5. Addition of homopolymer tails to DNA

Double-stranded DNA was tailed using terminal transferase with dCTP at 0.001 - 0.01 $\mu\text{g}/\mu\text{l}$ in the presence of 140 mM potassium cacodylate pH 7.6, 30 mM tris base, 0.1 mM DTT,

1 mM CoCl_2 and (^3H) or (^{32}P) - labelled dCTP in
75 - 150 fold excess over 3' termini. The reaction
was carried out at 37 C for 6 minutes. The extent
of incorporation of label was followed by assaying
5 the amount of TCA - insoluble radioactivity as a
proportion of the total radioactivity, counting
in Bray's scintillant.

The reaction was stopped by chilling and adding
EDTA to 10 mM, after which unincorporated material
10 was removed by gel filtration as described. Tailed
cDNA was precipitated as before, and dissolved in
1 M NaAc pH 8, 10 mM tris-acetate pH 8, 1 mM EDTA
ready for fractionation.

15

PstI-cleaved pBR322 DNA was similarly treated,
except that dGTP replaced dCTP.

6. Fractionation of tailed cDNA

cDNA was fractionated on 5 - 20% linear sucrose
20 density gradients in 1 M NaAc pH 8, 10 mM tris-
acetate pH 8, 1 mM EDTA, and centrifuged overnight
at 39,000 rpm in an SW50.1 rotor. DNA sedimentation
was checked on a parallel gradient loaded with a
mixture of HinfI and PstI digests of pBR322 DNA, and
25 fractions of this gradient were run on a 1% neutral
agarose gel. Fractions from the cDNA gradient

were diluted with an equal volume of water and precipitated with 2 volumes of cold ethanol, and then pooled to give three final fractions, a large cDNA fraction (larger than 2,200 bp), an intermediate fraction (1,000 - 2,200 bp) and a fraction containing smaller cDNAs (600 - 1,000 bp). cDNA molecules smaller than 600 bp were discarded.

The three final fractions were dissolved to approximately 5 ng/ μ l in 150 mM RbCl, 10 mM tris-HCl pH 7.6, 0.2 mM EDTA.

B. Annealing and transformation

1. Annealing

dC-tailed cDNA was mixed with dG-tailed pBR322 or pAT153 in approximately equimolar quantities, at a concentration of 0.4 ng/ μ l of vector. Buffers were as described above. The mixtures were heated to 70°C for 30 minutes and then cooled overnight to room temperature, and slowly chilled to 4°C. Competent cells were added and transformed as described below.

2. Preparation of competent cells and transformation

DH1 cells^{*} were grown in 10 ml cultures of psi broth (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 20 mM MgCl₂, pH adjusted to 7.6 with KOH; all bacteriological reagents from Difco), and grown at 37°C in a shaking waterbath to A₅₅₀ = 0.3. 1 ml of this was then inoculated into 25 ml of the

* (recA1, nalA, r_k⁻, m_k⁺, ndo1⁻, R⁻, relA1?)

same medium and grown to $A_{550} = 0.48$. The cells were then chilled on ice for 15 minutes and harvested at 5,000 rpm for 5 minutes in an MSE 21 centrifuge at 4°C. They were then resuspended in 10 ml of
5 100 mM RbCl, 50 mM $MgCl_2$, 10 mM $CaCl_2$, 35 mM NaAc pH 5.8, 15% glycerol and kept on ice for 15 minutes.

The cells were again harvested, and resuspended in 1 ml of 10 mM RbCl, 75 mM $CaCl_2$, 10 mM MOPS -
10 KOH pH 5.8, 15% glycerol, and kept on ice for a further 15 minutes.

100 μ l of the cells thus prepared were mixed with the annealed DNA samples, and incubated on ice for 30 minutes, after which they were heat-shocked
15 at 42°C for 90 - 120 seconds. 1 ml of psi broth was added, and the cells were grown at 37°C for 1 hour. They were then centrifuged briefly, and resuspended in 100 μ l of psi broth and plate on LB plates containing 14 μ g/ml tetracycline (LB is
20 1% tryptone, 0.5% yeast extract, 170 mM NaCl, 1.5% agar).

After 18 - 24 hours growth at 37°C, colonies were counted and spotted onto LB plates containing 33 μ g/ml ampicillin to identify those transformants
25 containing recircularised or uncut plasmids. Over 1600 Tet^r Amp^s clones were picked and transferred

in order d arrays onto large plates of LB containing 14 µg/ml tetracyclin . .

C. Screening

1. Labelling of oligonucleotide

5 The ricin B chain specific oligomer (20 mer)
was end labelled using polynucleotide kinase.
500 ng of oligonucleotide was incubated in 50 mM
tris pH 8.5, 10 mM MgCl₂, .5 mM DTT, 0.1 mM spermidine-
HCl, 0.1 mM EDTA with 60 µCi γ(³²P) ATP and 1 µl
10 polynucleotide kinase (Boehringer) for 35 min at
37°C. The reaction was stopped by adding an equal
volume of 0.6 M NH₄AC and the bulk of non incorporated
γATP was removed by passage through a sephadex G25
column in 0.14 M NaCl, 0.02 M tris pH 7.6, 0.005 M
15 EDTA, 0.1% SDS. The probe was stored frozen
at -20°C.

2. Colony Hybridization using oligonucleotide probe

Transformants were grown on nitrocellulose filters
(Schleicher & Schuell 0.45 µ) layered over LB plus
20 tetracycline. The filters, in triplicate, were
then transferred to LB-Tet plates containing 200
µg/ml chloramphenicol f r 16 h at 37°. The filters
were placed colony side up on two sheets of 3 mm
paper wetted with 0.5 M NaOH for 15 mins at room

temperature. The same procedure was followed for the following two washes (1) with 1M tris pH 8.0, and (2) with 1 M tris pH 8, 1.5 M NaCl (30 mins). The filters were air dried and baked at 80°C.

5 Prehybridizations and hybridizations were done in double sealed polythene bags. The filters were prehybridized in 0.9 M NaCl, 0.09 M tris 7.4, 0.006 M EDTA, 0.5% NP40, 2x Denhardts, 0.2% SDS, 100 µg/ml denatured single strand salmon sperm DNA and 70
10 µg/ml tRNA. Prehybridization was done for 4 h at 55°C. The prehybridization buffer was then squeezed from the bag and fresh buffer added that contained 50 ng labelled probe (to give a maximum concentration of 5 ng/ml buffer). Annealing was
15 done overnight at 37°C.

Non stringent washes were done with 6 x SSC at room temperature. The filters were washed in 4 changes of 6 x SSC over 3h. The triplicate filters were then washed at three different temperatures
20 determined from the base composition and degree of mismatch of the probe. Using 2°C for every A or T and 4°C for every C or G in the probe the wash temperatures selected were 52°C, 56°C and 60°C. The filters were washed at the stringent temperature in 6 x
25 SSC for 10 minutes and then dried thoroughly. The filters were exposed to X-ray film overnight.

D. Hybrid selection procedure

0145111

1. DNA binding

Plasmid DNA was purified from the positive clone(s) and 10-15 µg linearized with EcoRI. After phenol:
5 chloroform extraction and ethanol precipitation the pellet was dissolved in 0.5 ml 0.1 x SSC.
0.5 ml 1 M NaOH was then added and the mixture allowed to stand for 15 mins at room temperature. 4 ml
of a prechilled neutralizing solution (1.5 M NaCl,
10 0.25 M HCl, 0.25 M tris- (pH 8.0) was added and the 5 ml DNA sample sucked by vacuum through swinnies containing wetted Schleicher and Schuell 0.45 µ
filter discs. 5 ml 6 x SSC was then passed through the filter(s). These were air dried and then baked
15 for 2 h at 80°C.

2. Hybrid selection protocol

The filter(s) were placed in 5 ml bottles and prehybridized for 4 h at 41°C in 50% formamide,
0.4 M NaCl, 10 mM pipes-NaOH pH 6.4, 4 mM EDTA,
20 0.5 µg/ml tRNA, 10 µg/ml poly(A). The buffer was removed and the filter(s) typically hybridized overnight at 41°C in 50% formamide buffer (above) containing approx: 20 µg poly (A). RNA from castor beans. The buffer was removed and the filters washed twice for 15 mins in each
25 of the following: (1) 1 x SSC, 0.5% SDS at room

temperature, (2) 0.1 x SSC, 0.1% SDS at room temperature,
(3) 0.1 x SSC, 0.1% SDS at 50°C, (4) 0.1 x SSC, 0.1%
SDS at room temperature. The filters were drained
and 200 µl hybrid release buffer (90% formamide, 10
5 mM pipes-NaOH pH 6.4, 1 mM EDTA, 0.5% SDS) added to
each and mixed for 30 mins at 40°C. The buffer was
removed into a fresh eppendorf and NaCl added to .2 M.
The released mRNA was precipitated with ethanol, rinsed
several times in 70% ethanol, dried and dissolved in
10 5 µl sterile water. The sample was translated in a
reticulocyte lysate cell free system and the products
run out directly on an SDS-polyacrylamide gel or firstly
immunoprecipitated with appropriate antisera.

The DNA sequence coding for the above-mentioned
15 ricin precursor polypeptide of two of the above-mentioned
clones, referred to herein as pBRQ 6 and pBRCL 17 (RCL
= Ricinus communis lectin) has now been determined
by a combination of the Sanger dideoxy method (Sanger et
al, 1977 - Proc. Natl. Acad. Sci. U.S.A. 74, 5463-67) and
20 the procedure of Maxam and Gilbert (Maxam and Gilbert, 1980 .
Meth. Enzym. 65, 499-560). In order to determine the sequence
at the ends of each insert the inserts were excised from
pBR322 with Pst I and ligated into Pst I linearized,
phosphatased plasmid pUC6 (Vierra and Messing, 1982 -
25 Gene 19, 259-268). E. Coli DHI cells were transformed
by these recombinant plasmids. These new recombinant
plasmids are referred to herein as pRCL6 and pRCL17.

It is apparent 01 451 11

the two inserts contain a region of common sequence
and that together they represent a total ricin precursor
sequence. There are no nucleotide differences between
5 the overlapping regions of the inserts in pRCL6 and
pRCL17.

A new recombinant DNA molecule was then constructed that contains
the complete nucleotide sequence encoding the ricin
precursor polypeptide. This was achieved by isolating
10 a fragment, 323 base pairs in length, obtained from
pRCL17 by digestion with the restriction endonuclease
Sau 961, and ligating this fragment to a fragment
15 1561 base pairs in length isolated after a partial
digestion of pRCL6 with Sau 961. Ligation was performed
in 50 mM trisHCl (pH 7.4) 10 mM $MgCl_2$, 10 mM dithiothreitol,
1 mM spermidine, 10 mM ATP, 0.1 mg/ml BSA with 5 units
commercial T4 DNA ligase, and incubation proceeded
overnight at 15°C. After a standard phenol/chloroform
extraction and ethanol precipitation the ligated DNA
20 was pelleted, dissolved in a small volume of 10 mM
tris HCl (pH 7.4), 1mM EDTA and digested to completion
with Pst I. The resulting linearised DNA was then
ligated (as above) with an equal quantity of Pst I
linearised, phosphatased pUC8. The new recombinant DNA molecule
25 containing the entire DNA sequence of the ricin pre-
cursor and referred to as pRCL617, was used in
conventional manner to transform E. coli DHI cells.

The nucleotide sequence of pRCL617 is shown hereinafter.

This sequence was deduced from the two overlapping cDNA inserts in clones pRCL6 and pRCL17 (the limits of the DNA inserts in each of these two clones are given below).

Nucleotide residues are numbered in the 5' to 3' direction with the first residue of the codon specifying the amino terminal residue of mature ricin A chain numbered 1 and the nucleotides on the 5' side of residue 1 indicated by negative numbers. The 5' terminal sequence does not extend to the 5' end of the mRNA whereas the 3' terminal sequence shown is followed by a poly (dA) tract 27 residues long, thus representing the complete sequence of the region.

The predicted amino acid sequence is given below the nucleotide sequence and differences with the published amino acid sequence of mature ricin A and B chains (Funatsu G., Kimura, M and Funatsu, M. Agric.Biol. Chem. Vol 43, pp 2221-2224 (1979), and Yoshitake, S., Funatsu, G and Funatsu, M - Agric. Biol. Chem. Vol.42, pp 1267-1274 (1978)) are indicated underneath.

Residues absent from the published amino acid sequence are underlined with a dashed line and the position of amino acids present in the published sequence but absent from the derived sequence presented here are indicated by an asterisk. The dashed line beneath the

12 amino acid sequence linking the C-terminus of the
A chain and the N-terminus of the B chain is bracketed.
Amino acids are numbered from the amino terminal residue
of the mature A chain and the preceding residues are
5 indicated by negative numbers. Potential sites for
asparagine linked N-glycosylation are boxed and potential
poly (A) signals are underlined. The insert of pRCL6
extends from nucleotide - 102 to residue 1512 and the
insert of pRCL17 extends from nucleotide 733 to residue
10 1782.

The intervening twelve triplets code for the
linker amino acid sequence which is present in the pre-
cursor polypeptide and which is enzymatically removed
in the cell to separate the A and B chains. which chains
15 are joined by a disulphide bridge during the formation
of the ricin molecule itself. This linker region as
well as the presumptive amino terminal leader or signal
sequence (amino acids - 24 to -1) are not present in
the sequences already published by Funatsu et al.

20 Preproricin is the whole polypeptide coded for
by the aforesaid DNA insert, i.e. from amino acid -24
to amino acid 541. Proricin, which is obtained from
preproricin in the organism by removal of the amino
acid leader sequence, extends from amino acid 1 to amino
25 acid 541.

0145111

-100

-50

5'--AAACCGGGAG GAAATACTAT TGTAATATGG ATG TAT GCA GTG GCA ACA TGG CTT
Met Tyr Ala Val Ala Thr Trp Leu

-20

-1 1

TGT TTT GGA TCC ACC TCA GGG TGG TCT TTC ACA TTA GAG GAT AAC AAC ATA
Cys Phe Gly Ser Thr Ser Gly Trp Ser Phe Thr Leu Glu Asp Asn Asn Ile
-10 -1 1

50

TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGT GCC ACT GTG
Phe Pro Lys Gln Tyr Pro Ile Ile Asn Phe Thr Thr Ala Gly Ala Thr Val
10

100

CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TTA ACA ACT GGA
Gln Ser Tyr Thr Asn Phe Ile Arg Ala Val Arg Gly Arg Leu Thr Thr Gly
20 30

150

GCT GAT GTG AGA CAT GAT ATA CCA GTG TTG CCA AAC AGA GTT GGT TTG CCT
Ala Asp Val Arg His Asp Ile Pro Val Leu Pro Asn Arg Val Gly Leu Pro
40 Glu 50

200

ATA AAC CAA CGG TTT ATT TTA GTT GAA CTC TCA AAT CAT GCA GAG CTT TCT
Ile Asn Gln Arg Phe Ile Leu Val Glu Leu Ser Asn His Ala Glu Leu Ser
60 Gln

250

GTT ACA TTA GCC CTG GAT GTC ACC AAT GCA TAT GTG GTC GGC TAC CGT GCT
Val Thr Leu Ala Leu Asp Val Thr Asn Ala Tyr Val Val Gly Tyr Arg Ala
70 Ser 80

300

GGA AAT AGC GCA TAT TTC TTT CAT CCT GAC AAT CAG GAA GAT GCA GAA GCA
Gly Asn Ser Ala Tyr Phe Phe His Pro Asp Asn Gln Glu Asp Ala Glu Ala
90 100

350

ATC ACT CAT CTT TTC ACT GAT GTT CAA AAT CGA TAT ACA TTC GCC TTT GGT
Ile Thr His Leu Phe Thr Asp Val Gln Asn Arg Tyr Thr Phe Ala Phe Gly
110 120

400

GGT AAT TAT GAT AGA CTT GAA CAA CTT GCT GGT AAT CTG AGA GAA AAT ATC
Gly Asn Tyr Asp Arg Leu Glu Gln Leu Ala Gly Asn Leu Arg Glu Asn Ile
130

450

GAG TTG GGA AAT GGT CCA CTA GAG GAG GCT ATC TCA GCG CTT TAT TAT TAC
Glu Leu Gly Asn Gly Pro Leu Glu Glu Ala Ile Ser Ala Leu Tyr Tyr Tyr
140 150

500

AGT ACT GGT GGC ACT CAG CTT CCA ACT CTG GCT CGT TCC TTT ATA ATT TGC
Ser Thr Gly Gly Thr Gln Leu Pro Thr Leu Ala Arg Ser Phe Ile Ile Cys
160 170

550

ATC CAA ATG ATT TCA GAA GCA GCA AGA TTC CAA TAT ATT GAG GGA GAA ATG
Ile Gln Met Ile Ser Glu Ala Ala Arg Phe Gln Tyr Ile Glu Gly Glu Met
180

600

CGC ACG AGA ATT AGG TAC AAC CGG AGA TCT GCA CCA GAT CCI AGC GTA ATT
Arg Thr Arg Ile Arg Tyr Asn Arg Arg Ser Ala Pro Asp Pro Ser Val Ile
190 200

650

ACA CTT GAG AAT AGT TGG GGG AGA CTT TCC ACT GCA ATT CAA GAG TCT AAC
Thr Leu Glu Asn Ser Trp Gly Arg Leu Ser Thr Ala Ile Gln Glu Ser Asn
210 220

700

CAA GGA GCC TTT GCT AGT CCA ATT CAA CTG CAA AGA CGT AAT GGT TCC AAA
Gln Gly Ala Phe Ala Ser Pro Ile Gln Leu Gln Arg Arg Asn Gly Ser Lys
230 --- Asp

750

TTC AGT GTG TAC GAT GTG AGT ATA TTA ATC CCT ATC ATA GCT CTC ATG GTG
Phe Ser Val Tyr Asp Val Ser Ile Leu Ile Pro Ile Ile Ala Leu Met Val
240 Leu 250 ---

800

TAT AGA TGC GCA CCT CCA CCA TCG TCA CAG TTT TCT TTG CTT ATA AGG CCA
Tyr Arg Cys Ala Pro Pro Ser Ser Gln Phe Ser Leu Leu Ile Arg Pro
260 (--- --- --- --- ---)

GTG GTA CCA AAT TTT AAT GCT GAT GTT TGT ATG GAT CCT GAG CCC ATA GTG
Val Val Pro Asn Phe Asn Ala Asp Val Cys Met Asp Pro Glu Pro Ile Val
--- --- --- --- --- ---)280 290

900

CGT ATC GTA GGT CGA AAT GGT CTA TGT GTT GAT GTT AGG GAT GGA AGA TTC
Arg Ile Val Gly Arg Asn Gly Leu Cys Val Asp Val Arg Asp Gly Arg Phe
300 Asn

950

CAC AAC GGA AAC GCA ATA CAG TTG TGG CCA TGC AAG TCT AAT ACA GAT GCA
His Asn Gly Asn Ala Ile Gln Leu Trp Pro Cys Lys Ser Asn Thr Asp Ala
Asn His 310 320

1000

AAT CAG CTC TGG ACT TTG AAA AGA GAC AAT ACT ATT CGA TCT AAT GGA AAG
Asn Gln Leu Trp Thr Leu Lys Arg Asp Asn Thr Ile Arg Ser Asn Gly Lys
--- 330 340

1050

TGT TTA ACT ACT TAC GGG TAC AGT CCG GGA GTC TAT GTG ATG ATC TAT GAT
Cys Leu Thr Thr Tyr Gly Tyr Ser Pro Gly Val Tyr Val Met Ile Tyr Asp
Pro Ser

1100

TGC AAT ACT GCT GCA ACT GAT GCC ACC CGC TGG CAA ATA TGG GAT
Cys Asn Thr Ala Ala Thr Asp Ala Thr Arg Trp Gln Ile Trp Asp
360 Thr Asp --- Glu Asn

1150

AAT GGA ACC ATC ATA AAT CCC AGA TCT AGT CTA GTT TTA GCA GCG ACA TCA
Asn Gly Thr Ile Ile Asn Pro Arg Ser Ser Leu Val Leu Ala Ala Thr Ser
380 390

1200

GGG AAC AGT GGT ACC ACA CTT ACG GTG CAA ACC AAC ATT TAT GCC GTT AGT
Gly Asn Ser Gly Thr Thr Leu Thr Val Gln Thr Asn Ile Tyr Ala Val Ser
400

1250

CAA GGT TGG CTT CCT ACT AAT AAT ACA CAA CCT TTT GTT ACA ACC ATT GTT
Gln Gly Trp Leu Pro Thr Asn Asn Thr Gln Pro Phe Val Thr Thr Ile Val
Pro Phe Trp 420

Page 30c follows...

1300

GGG CTA TAT GGT CTG TGC TTG CAA GCA AAT AGT GGA CAA GTA TGG ATA GAG
Gly Leu Tyr Gly Leu Cys Leu Gln Ala Asn Ser Gly Gln Val Trp Ile Glu
430 Val 440

1350

GAC TGT AGC AGT GAA AAG GCT GAA CAA CAG TGG GCT CTT TAT GCA GAT GGT
Asp Cys Ser Ser Glu Lys Ala Glu Gln Gln Trp Ala Leu Tyr Ala Asp Gly
Ser Cys 450 Ser

1400

TCA ATA CGT CCT CAG CAA AAC CGA GAT AAT TGC CTT ACA AGT GAT TCT AAT
Ser Ile Arg Pro Gln Gln Asn Arg Asp Asn Cys Leu Thr Ser Asp Ser Asn
Asn 460 Asn Arg 470

1450

ATA CGG GAA ACA GTT GTT AAG ATC CTC TCT TGT GGC CCT GCA TCC TCT GGC
Ile Arg Glu Thr Val Val Lys Ile Leu Ser Cys Gly Pro Ala Ser Ser Gly
480 490

1500

CAA CGA TGG ATG TTC AAG AAT GAT GGA ACC ATT TTA AAT TTG TAT AGT GGA
Gln Arg Trp Met Phe Lys Asn Asp Gly Thr Ile Leu Asn Leu Tyr Ser Gly
500

1550

TTG GTG TTA GAT GTG AGG CGA TCG GAT CCG AGC CTT AAA CAA ATC ATT CTT
Leu Val Leu Asp Val Arg Arg Ser Asp Pro Ser Leu Lys Gln Ile Ile Leu
510 Ala 520

1600

TAC CCT CTC CAT GGT GAC CCA AAC CAA ATA TGG TTA CCA TTA TTT TGA
Tyr Pro Leu His Gly Asp Pro Asn Gln Ile Trp Leu Pro Leu Phe
Trp * --- --- *Leu Pro

1650

TAGACAGATT ACTCTCTTGC AGTGTGTGTG TCCTGCCATG AAAATAGATG GCTTAAATAA

1700

AAAGGACATT GTAAATTTTG TAACTGAAAG GACAGCAAGT TATTGCAGTC CAGTATCTAA

1750

1780

TAAGAGCACA ACTATTGTCT TGTGCATTCT AAATTT-Poly(A)

CLAIMS:

1. DNA comprising a nucleotide sequence coding for at least a substantial portion of a plant toxin of the ricin-type or a mutant thereof, characterised in that it is biologically pure and homogeneous.
2. DNA according to claim 1, characterised in that the nucleotide sequence codes for the A chain or the B chain of the mature toxin.
3. DNA according to claim 1, 2 or 3, characterised in that the nucleotide sequence codes for a mutant in which the galactose binding site or sites have been removed or inactivated.
4. A sample of DNA including at least a substantial portion of the following nucleotide sequence:

0145111

ATG TAT GCA GTG GCA ACA TGG CTT

TGT TTT GGA TCC ACC TCA GGG TGG TCT TTC ACA TTA GAG GAT AAC AAC ATA

TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGT GCC ACT GTG

CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TTA ACA ACT GGA

5. GCT GAT GTG AGA CAT GAT ATA CCA GTG TTG CCA AAC AGA GTT GGT TTG CCT

ATA AAC CAA CGG TTT ATT TTA GTT GAA CTC TCA AAT CAT GCA GAG CTT TCT

GTT ACA TTA GCC CTG GAT GTC ACC AAT GCA TAT GTG GTC GGC TAC CGT GCT

GGA AAT AGC GCA TAT TTC TTT CAT CCT GAC AAT CAG GAA GAT GCA GAA GCA

ATC ACT CAT CTT TTC ACT GAT GTT CAA AAT CGA TAT ACA TTC GCC TTT GGT

Page 32a follows

0145111

GGI AAT TAT GAT AGA CTT GAA CAA CTI GCI GGT AAT CTG AGA GAA AAT ATC

GAG ITG GGA AAT GGT CCA CTA GAG GAG GCI ATC TCA GCG CTT TAT TAT TAC

AGT ACT GGI GGC ACT CAG CTT CCA ACT CTG GCI CGI TCC TTT ATA ATT TGC

ATC CAA ATG ATT TCA GAA GCA GCA AGA TTC CAA TAT ATT GAG GGA GAA ATG

5. CGC ACG AGA ATT AGG TAC AAC CGG AGA TCI GCA CCA GAT CCI AGC GTA ATT

ACA CTI GAG AAT AGI TGG GGG AGA CTI TCC ACT GCA ATT CAA GAG TCI AAC

CAA GGA GCC TTI GCI AGT CCA ATT CAA CTG CAA AGA CGI AAT GGI TCC AAA

TTC AGI GTG TAC GAT GTG AGI AIA TTA AIC CCI ATC AIA GCT CTC ATG GTG

TAT AGA TGC GCA CCI CCA CCA TCG TCA CAG TIT TCI TTG CTI AIA AGG CCA

Page 32b follows...

0145111

GJG GJA CCA AAT JJJ AAT GCT GAT GJJ JGJ ATG GAT CCT GAG CCC ATA GJG

--

CGJ AJC GJA GGT CGA AAT GGT CJA TGT GJJ GAT GTJ AGG GAT GGA AGA TJC

CAC AAC GGA AAC GCA ATA CAG TTG TGG CCA TGC AAG TCT AAT ACA GAT GCA

AAT CAG CTG TGG ACT TTG AAA AGA GAC AAT ACT ATT CGA TCT AAT GGA AAG

5. TGT TJA ACT ACT TAC GGG TAC AGT CCG GGA GTC TAT GJG ATG AJC TAT GAT

TGT AAT ACT GCT GCA ACT GAT GGC ACC CGT TGS CAA ATA TGS GAT

AAT GGA ACC ATC ATA AAT CCC AGA TCT AGT CJA GTJ TJA GCA GCG ACA TCA

GGS AAC AGT GGT ACC ACA CTJ ACG GJG CAA ACC AAC ATT TAT GCC GTJ AGT

CAA GGT TGG CTJ CCT ACT AAT AAT ACA CAA CCT TJJ GTJ ACA ACC ATT GTJ

Page 32c follows...

0145111

GGG CTA TAT GGT CTG TGC TTG CAA GCA AAT AGT GGA CAA GAA TGG ATA GAG

GAC TGT AGC AGT GAA AAG GCT GAA CAA CAG TGG GCT CTT TAT GCA GAT GGT

TCA ATA CGT CCT CAG CAA AAC CGA GAT AAT TGC CTT ACA AGT GAT TCT AAT

ATA CGG GAA ACA GTT GTT AAG ATC CTC TCT TGT GGC CCT GCA TCC TCT GGC

5 CAA CGA TGG ATG TTC AAG AAT GAT GGA ACC ATT TTA AAT TTG TAT AGT GGA

TTG GTG TTA GAT GTG AGG CGA TCG GAT CCG AGC CTT AAA CAA ATC ATT CTT

TAC CCT CTC CAT GGT GAC CCA AAC CAA ATA TGG TTA CCA TTA TTT

or at least a portion of a nucleotide sequence that is
equivalent thereto by virtue of degeneracy of the genetic
10 code.

5. A sample of DNA including at least a substantial portion
of the following nucleotide sequence:

ATG TAT GCA GTG GCA ACA TGG CTI

TGI TTI GGA TCC ACC TCA GGG TGG TCI TTC ACA TTA GAG GAT AAC AAC ATA

TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGT GGC ACT GIG

CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TTA ACA ACT GGA

5 GCT GAT GTG AGA CAT GAT ATA CCA GTG TTG CCA AAC AGA GTT GGT TIG CCI

ATA AAC CAA CGG TTT ATT TTA GTT GAA CTC TCA AAT CAT GCA GAG CTI TCI

GTT ACA TTA GTC CTG GAT GTC ACC AAT GCA TAT GTG GTC GGC TAC CGI GCI

GGA AAT AGC GCA TAT TTC TTT CAT CCI GAC AAT CAG GAA GAT GCA GAA GCA

ATC ACT CAT CTI TTC ACT GAT GTT CAA AAT CGA TAT ACA TTC GCC TTT GGT

Page 33a follows...

GGT AAT TAT GAT AGA CTT GAA CAA CTT GCT GGT AAT CTG AGA GAA AAT ATC

GAG TTG GGA AAT GGT CCA CTA GAG GAG GCT ATC TCA GCG CTT TAT TAT TAC

AGT ACT GGT GGC ACT CAG CTT CCA ACT CTG GCT CGT TCC TTT ATA ATT TGC

ATC CAA ATG ATT TCA GAA GCA GCA AGA TTC CAA TAT ATT GAG GGA GAA ATG

5 CGC ACG AGA ATT AGG TAC AAC CGC AGA TCT GCA CCA GAT CCT AGC GTA ATT

ACA CTG GAG AAT AGT TGG GGG AGA CTT TCC ACT GCA ATT CAA GAG TCT AAC

CAA GGA GCC TTT GCT AGT CCA ATT CAA CTG CAA AGA CGT AAT GGT TCC AAA

TTT AGT GTG TAC GAT GTG AGT ATA TTA ATC CCT ATC ATA GCT CTC ATG GTG

TAT AGA TGC GCA CCT CCA CCA TCG TCA CAG TTT ,

10 or at least a portion of a nucleotide sequence which is
equivalent thereto by virtue of degeneracy of the genetic
code.

0145111

6. A sample of DNA including at least a substantial portion of the following nucleotide sequence:

GCT GAT GTT TGT ATG GAT CCT GAG CCC ATA GTG

CGT ATC GTA GGT CGA AAT GGT CTA TGT GTT GAT GTT AGG GAT GGA AGA TTC

5 CAC AAC GGA AAC GCA ATA CAG TTG TGG CCA TGC AAG TCT AAT ACA GAT GCA

AAT CAG CTT TGG ACT TTG AAA AGA GAC AAT ACT ATT CGA TCT AAT GGA AAG

TGT TTA ACT ACT TAC GGG TAC AGT CCG GGA GTC TAT GTG ATG ATC TAT GAT

TGC AAT ACT GCT GCA ACT GAT GCC ACC CGC TGG CAA ATA TGG GAT

AAT GGA ACC ATC ATA AAT CCC AGA TCT AGT CTA GTT TTA GCA GCG ACA TCA

10 GGG AAC AGT GGT ACC ACA CTT ACG GTG CAA ACC AAC ATT TAT GCC GTT AGT

CAA GGT TGG CTT CCT ACT AAT AAT ACA CAA CCT TTT GTT ACA ACC ATT GTT

Page 34a follows...

GGG CTA TAT GGT CTG TGC TTG CAA GCA AAT AGT GGA CAA GAA TGG ATA GAG

GAC TGT AGC AGT GAA AAG GCT GAA CAA CAG TGG GCT CTT TAT GCA GAT GGT

TCA ATA CGT CCT CAG CAA AAC CGA GAT AAT TGC CTT ACA AGT GAT TCT AAT

ATA CGG GAA ACA GTT GTT AAG ATC CTC TCT TGT GGC CCT GCA TCC TCT GGC

5 CAA CGA TGG ATG TTC AAG AAT GAT GGA ACC ATT TTA AAT TTG TAT AGT GGA

TTG GTG TTA GAT GTG AGG CGA TCG GAT CCG AGC CTT AAA CAA ATC ATT CTT

TAC CCI CTC CAT GGT GAC CCA AAC CAA ATA TGG TTA CCA TTA TTI

or at least a portion of a nucleotide sequence which is
equivalent thereto by virtue of degeneracy of the genetic
10 code.

7. A recombinant DNA molecule, characterised in that it
includes as an insert a nucleotide sequence as defined in
any preceding claim.

8. A recombinant DNA molecule according to claim 7,
15 characterised in that said insert has been introduced

into a cloning vector which is a plasmid or a bacteriophage.

9. A modified host cell containing a recombinant DNA molecule according to claim 7 or claim 8.

5 10. A modified host cell according to claim 9, characterised in that the host is a plant cell, an animal cell, a Gram-negative bacterium, a Gram-positive bacterium, or a yeast.

10 11. A modified host microorganism according to claim 10, characterised in that the host is a Gram-negative bacterium, which is any of E.coli, Methylophilus methylotropus, or Alcaligenes eutrophus.

15 12. A modified host microorganism according to claim 10, characterised in that the host is a Gram-positive bacterium which is any of Streptomyces, Bacillus or Arthrobacter.

13. A modified host microorganism according to claim 10, characterised in that the host is a yeast which is Saccharomyces cerevisiae.

20 14. A modified host microorganism according to

claim 11, wherein the cloning vector is any suitable plasmid selected from pBR322, pAT153, pUC8, pGS15 or pMB9.

15. A modified host microorganism according to
5 claim 12, wherein the cloning vector is the plasmid pUC6 .

16. A modified host microorganism according to claim 13, when the cloning vector is pMA91, pMA230, YRp7, pLC544, and YEp6.

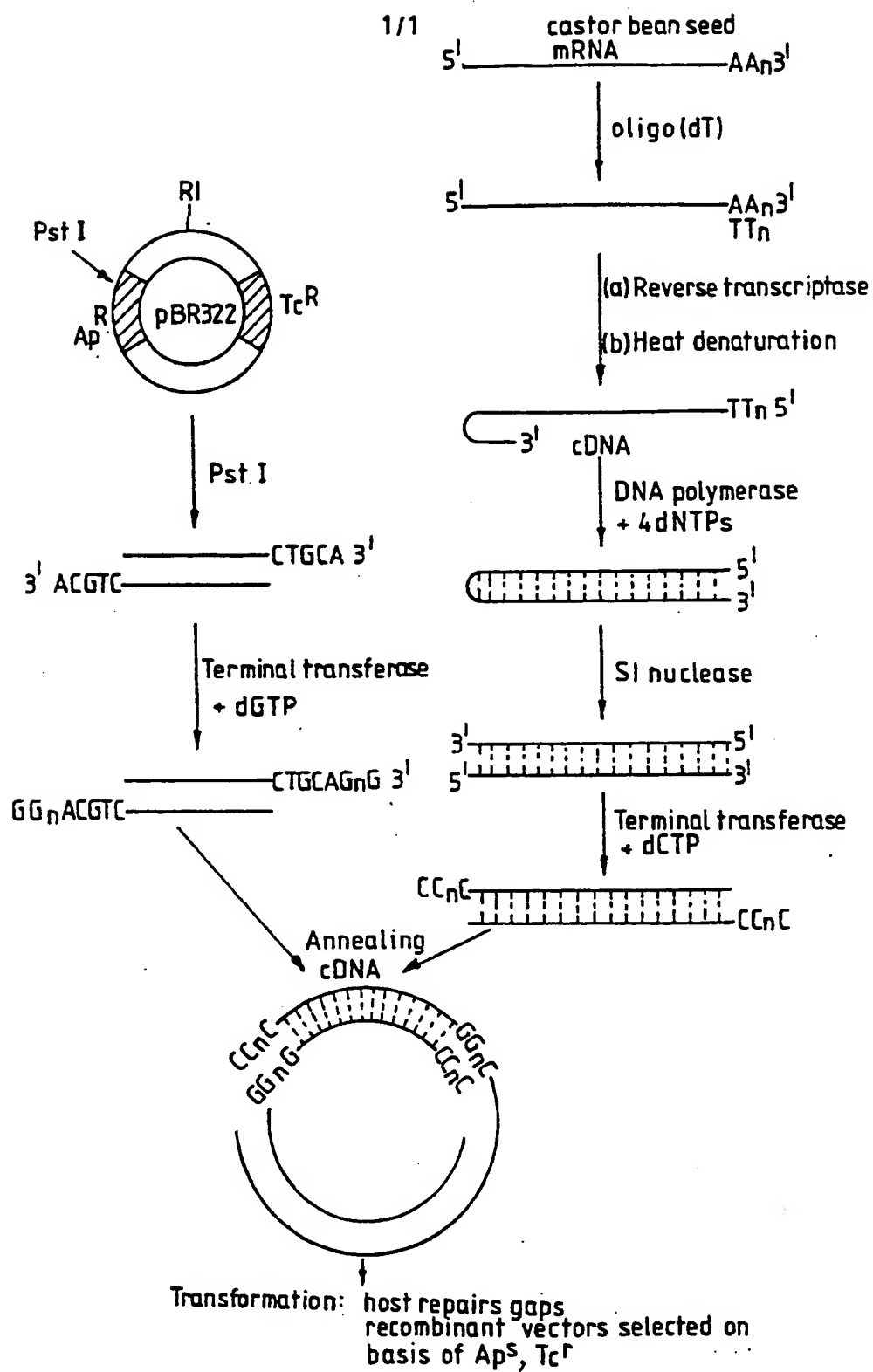
10 17. Method of preparing a biologically-pure sample of a cDNA sequence coding for a precursor, or a portion thereof, of a plant toxin of the ricin type, characterised by isolating mRNA from the tissue of a plant which produces such a toxin, and synthesising
15 cDNA from this by reverse transcription.

18. A method of obtaining a recombinant DNA molecule by inserting a double-stranded cDNA according to claim 17 into a cloning vector.

19. A method of obtaining a genetically-modified
20 host characterised in that a recombinant DNA molecule according to claim 18 is introduced into a host micro-organism.

4 Dyer's Buildings, Holborn, London EC1N 2JT.

- See: State Police Department, Vol. 3, 1964, in: 1964, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642,





DOCUMENTS CONSIDERED TO BE RELEVANT			EP 84304801.8
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl. 4)
A,D	<p>AGRICULTURAL AND BIOLOGICAL CHEMISTRY, vol. 42, no. 5, May 1978, Tokyo</p> <p>S. YOSHITAKE et al. "Isolation and Sequences of Peptic Peptides, and the Complete Sequence of Ile Chain of Ricin D" pages 1267-1274</p> <p>* Pages 1270-1272 *</p> <p>--</p>	1,2	<p>C 12 N 15/00</p> <p>C 12 P 21/00</p> <p>//C 12 R 1:19</p> <p>C 12 R 1:05</p> <p>C 12 R 1:465</p> <p>C 12 R 1:07</p> <p>C 12 R 1:06</p> <p>C 12 R 1:865</p>
A,D	<p>AGRICULTURAL AND BIOLOGICAL CHEMISTRY, vol. 43, no. 9, September 1979, Tokyo</p> <p>G. FUNATSU et al. "Primary Structure of Ala Chain of Ricin D" pages 2221-2224</p> <p>* Totality *</p> <p>----</p>	1,2	<p>TECHNICAL FIELDS SEARCHED (Int. Cl. 4)</p> <p>C 12 N</p> <p>C 12 P</p>
The present search report has been drawn up for all claims			
Place of search VIENNA		Date of completion of the search 19-10-1984	Examiner WOLF
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone</p> <p>Y : particularly relevant if combined with another document of the same category</p> <p>A : technological background</p> <p>O : non-written disclosure</p> <p>P : intermediate document</p> <p>T : theory or principle underlying the invention</p> <p>E : earlier patent document, but published on, or after the filing date</p> <p>D : document cited in the application</p> <p>L : document cited for other reasons</p> <p>& : member of the same patent family, corresponding document</p>			

12

EUROPEAN PATENT SPECIFICATION

45 Date of publication of patent specification: 09.05.90

21 Application number: 84304801.8

22 Date of filing: 13.07.84

6 Int. Cl.⁵: **C 12 N 15/00, C 12 P 21/00 //**
C12R1/19, C12R1/05,
C12R1/465, C12R1/07,
C12R1/06, C12R1/865

64 DNA sequence coding for a plant toxin of the ricin type, or a portion thereof.

30 Priority: 15.07.83 GB 8319265
13.03.84 GB 8406569

43 Date of publication of application:
19.06.85 Bulletin 85/25

45 Publication of the grant of the patent:
09.05.90 Bulletin 90/19

64 Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

56 References cited:

Wo 85103508

Eur.J. Biochem. 119, 31-41 (1981)

Eur.J. Biochem. 137, 57-65 (1983)

70 Proprietor: The University of Warwick
Coventry West Midlands, CV4 7AL (GB)

72 Inventor: Lord, John Michael
93 Campion Court
Leamington Spa Warwickshire (GB)
Inventor: Roberts, Lynne Margaret
48 Braemar Road Lillington
Leamington Spa Warwickshire (GB)
Inventor: Lamb, Francis Ian
76 Henry Street
Kenilworth Warwickshire CV8 2HJ (GB)

74 Representative: Bannerman, David Gardner
Withers & Rogers 4 Dyer's Buildings Holborn
London, EC1N 2JT (GB)

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European patent convention).

Courier Press, Leamington Spa, England.

EP 0 145 111 B1

Description

This invention relates to DNA comprising a nucleotide sequence coding for at least a portion of a polypeptide which is a plant toxin of the ricin type, as hereinafter defined. It also relates to recombinant DNA molecules comprising a DNA sequence which codes for a polypeptide which is closely related to a natural plant toxin of the ricin type. Ricin, and also other plant toxins such as abrin, modeccin and viscumin, consist of two polypeptide chains (known as the A and B chains) linked by a disulphide bridge, one chain (the A chain) being primarily responsible for the cytotoxic and the other chain (the B chain) having sites enabling the molecule to bind to cell surfaces. Ricin is produced in the plant *Ricinus communis* (also known as the castor bean plant) via a precursor protein known as "preproricin".

Preproricin comprises a single polypeptide chain which includes a leader sequence. The leader sequence is subsequently removed in the organism to give proricin which is then cleaved to eliminate a linker region and joined by a disulphide bond to form the mature protein.

The toxicity of ricin-type toxins operates in three phases: (1) binding to the cell surface via the B chain; (2) penetration of at least the A chain into the cytosol, and (3) inhibition of protein synthesis through the A chain attacking the 60S subunits of the ribosomes. Thus, separated A and B chains are essentially non-toxic, the inherently toxic A chain lacking the ability to bind to cell surfaces in the absence of the B chain.

It is also known that in ricin-type toxins the B chain binds to cell surfaces by virtue of galactose recognition sites, which react with glycoproteins or glycolipids exposed at the cell surface.

It has already been suggested that the toxicity of the ricin A chain might be exploited in anti-tumour therapy, by replacing the indiscriminately-binding B chain with a different carrier component having the ability to bind only to tumour cells. Thus, various immunotoxins have already been prepared, consisting of a conjugate of whole ricin or a separated natural ricin A chain and a tumour-specific monoclonal antibody. Although these known conjugates are of considerable potential in themselves, there is scope for improvement.

One problem with the known conjugates arises from a structural feature of the A chain from natural ricin. It is known that the natural ricin A chain becomes N-glycosylated during its synthesis, by enzymes present in *Ricinus* cells, and it is thought that the resulting sugar moieties are capable of non-specific interactions with cell surfaces. Thus, it appears that the known A chain conjugates are capable of a certain amount of binding with non target cells, even in the absence of the natural B chain, thus increasing the toxicity of such immunotoxins towards non target cells.

Another problem with the known ricin A chain conjugates stems from the fact that the B chain seems to have an important secondary function in that it somehow assists in the intoxication process, apart from its primary function in binding the ricin molecule to the cell surface. This secondary function is lost if the B chain is replaced by a different carrier component such as a monoclonal antibody.

If it were possible to prevent interactions between the cell surface via the A chain sugar moieties, whilst preserving the secondary toxicity-increasing function of the B chain, the toxicity of a whole ricin antibody conjugate towards normal cells could be reduced, and towards target cells could be increased, thus improving the therapeutic index of the immunotoxin. It is also known that the natural ricin B chain is N-glycosylated and the B chain sugar moieties may also contribute to non specific interactions. Also, the sugar moieties in both chains enable the ricin molecule to be sequestered by reticuloendothelial cells in the liver, and so would lead to the rapid excretion from the system of a drug based on a part or the whole of the ricin molecule in which such sugar moieties were still present.

Attempts to remove all the sugar moieties from natural ricin by chemical or enzymatic methods have so far failed. Nevertheless the major obstacle confronting the use of known whole ricin-antibody conjugates is the presence of two galactose binding sites in the ricin B chain. These B chain galactose binding sites are primarily responsible for the non-specific cellular interactions of current whole ricin-antibody conjugates, particularly when used *in vivo*. Their presence in the natural toxin clearly eliminates or reduces the targeting specificity conferred by the antibody.

An improved immunotoxin based on ricin or another plant toxin of the ricin type, not suffering from these problems, could consist of a whole toxin molecule modified so that it is not N-glycosylated, and so that the B chain has no galactose recognition sites, but retains its secondary intoxication-promoting properties, coupled to a carrier moiety which delivers the toxin to the target cells. This could be a tumour-specific or cell/tissue specific vehicle such as a suitable monoclonal antibody.

Our research which has so far been concentrated on ricin itself, has indicated that the assembly of ricin (and the related agglutinin which consists of two ricin-like molecules with slightly modified A and B chains) does not involve the separate synthesis of the A and B chains as the products of distinct mRNA's, but rather the initial formation of a single polypeptide precursor containing both the A chain and B chain sequences. This is thought to apply in the case of other toxins of the same type.

This invention is based on the idea of preparing a genetically-engineered microorganism capable of expressing a molecule of a toxin of the ricin type, as defined above, or alternatively part of such a toxin molecule, or a precursor of such a molecule (which could be converted to the toxin molecule itself) which toxin molecule could be modified as suggested above and could be used to construct an effective toxin conjugate by combining it with a tumour-specific or cell/tissue specific monoclonal antibody or other carrier moiety, such as a hormone or lectin.

EP 0 145 111 B1

The fact that ricin is formed via a precursor polypeptide will enable a cell system to be constructed by known techniques which expresses a ricin precursor. The ricin precursor product could then be chemically or enzymatically converted to the desired modified ricin. An analogous technique could be used in the case of other ricin-type toxins as herein defined. An alternative technique would be to divide from the DNA sequence that codes for the precursor two sequences which code separately for the A and B chains, to insert these separated sequences into different cloning vehicles and to insert the resulting recombinant DNA molecules into separate host microorganisms. One such host would then express the A chain polypeptide sequence and the other the B chain polypeptide sequence. These sequences could then be combined to form the desired modified ricin molecule. This technique could obviously also be used for any ricin-type toxins in which the A and B chains are encoded by distinct mRNA gene pools. This latter approach would be preferred on safety grounds, in that separate and therefore non-toxic A and B chains would be expressed.

According to one aspect of the invention we provide a biologically pure and homogeneous sample of DNA comprising a nucleotide sequence coding for at least a portion of a precursor of a ricin-type toxin polypeptide, or mutants thereof said portion coding for a polypeptide that displays the ribosome-inactivity property of the A chain of ricin.

Said portion preferably comprises the A chain of the mature protein.

More specifically, we provide a sample of DNA including at least a substantial portion of any of the following DNA sequences, said portion coding for a polypeptide that displays the ribosome-inactivity property of the A chain of ricin, which sample is biologically-pure:

ATG TAT GCA GTG GCA ACA TGG CTT

TGT TTT GGA TCC ACC TCA GGG TGG TCT TTC ACA TTA GAG GAT AAC AAC ATA

TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGT GCC ACT GTG

CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TTA ACA ACT GGA

GCT GAT GTG AGA CAT GAT ATA CCA GTG TIG CCA AAC AGA GTT GGT TTG CCT

ATA AAC CAA CGG TTT ATT TTA GTT GAA CTC TCA AAT CAT GCA GAG CTT TCT

GTG ACA TTA GCC CTG GAT GTC ACC AAT GCA TAT GTG GTC GGC TAC CGT GCT

GGA AAT AGC GCA TAT TTC TTT CAT CCT GAC AAT CAG GAA GAT GCA GAA GCA

ATC ACT CAT CTT TTC ACT GAT GTT CAA AAT CGA TAT ACA TTC GCC TTT GGT

GGT AAT TAT GAT AGA CTT GAA CAA CTT GCT GGT AAT CTG AGA GAA AAT ATC

GAG TTG GGA AAT GGT CCA CTA GAG GAG GCT ATC TCA GCG CTT TAT TAT TAC

AGT ACT GGT GGC ACT CAG CTT CCA ACT CTG GCT CGT TCC TTT ATA ATT TGC

ATC CAA ATG ATT TCA GAA GCA GCA AGA TTC CAA TAT ATT GAG GGA GAA ATG

CGC ACG AGA ATT AGG TAC AAC CGG AGA TCT GCA CCA GAT CCT AGC GTA ATT

ACA CTT GAG AAT AGT TGG GGG AGA CTT TCC ACT GCA ATT CAA GAG TCT AAC

EP 0 145 111 B1

CAA GGA GCC TTT GGT AGT CCA ATT CAA CTG CAA AGA CGT AAT GGT TCC AAA
TTC AGT GTG TAC GAT GTG AGT ATA TTA ATC CCT ATC ATA GCT CTC ATG GTG
TAT AGA TGC GCA CCT CCA CCA TCG TCA CAG TTT TCT TTG CTT ATA AGG CCA
GTG GTA CCA AAT TTT AAT GCT GAT GTT TGT ATG GAT CCT GAG CCC ATA GTG
CGT ATC GTA GGT CGA AAT GGT CTA TGT GTT GAT GTT AGG GAT GGA AGA TTC
CAC AAC GGA AAC GCA ATA CAG TTG TGG CCA TGC AAG TCT AAT ACA GAT GCA
AAT CAG CTC TGG ACT TTG AAA AGA GAC AAT ACT ATT CGA TCT AAT GGA AAG
TGT TTA ACT ACT TAC GGG TAC AGT CCG GGA GTC TAT GTG ATG ATC TAT GAT
TGC AAT ACT GCT GCA ACT GAT GCC ACC CGC TGG CAA ATA TGG GAT
AAT GGA ACC ATC ATA AAT CCC AGA TCT AGT CTA GTT TTA GCA GCG ACA TCA
GGG AAC AGT GGT ACC ACA CTT ACG GTG CAA ACC AAC ATT TAT GCC GTT AGT
CAA GGT TGG CIT CCT ACT AAT AAT ACA CAA CCT TTT GTT ACA ACC ATT GTT
GGG CTA TAT GGT CTG TGC TTG CAA GCA AAT AGT GGA CAA GTA TGG ATA GAG
GAC TGT AGC AGT GAA AAG GCT GAA CAA CAG TGG GCT CTT TAT GCA GAT GGT
TCA ATA CGT CCT CAG CAA AAC CGA GAT AAT TGC CTT ACA AGT GAT TCT AAT
ATA CGG GAA ACA GTT GTT AAG ATC CTC TCT TGT GGC CCT GCA TCC TCT GGC
CAA CGA TGG ATG TTC AAG AAT GAT GGA ACC ATT TTA AAT TTG TAT AGT GGA
TTG GTG TTA GAT GTG AGG CGA TCG GAT CCG AGC CTT AAA CAA ATC ATT CTT
TAC CCT CTC CAT GGT GAC CCA AAC CAA ATA TGG TTA CCA TTA TTT
ATG TAT GCA GTG GCA ACA TGG CTT
TGT TIT GGA TCC ACC TCA GGG TGG TCT TTC ACA TTA GAG GAT AAC AAC ATA
TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGI GCC ACT GTG

EP 0 145 111 B1

CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TIA ACA ACT GGA
GCT GAT GTG AGA CAT GAT ATA CCA GTG TIG CCA AAC AGA GTT GGT TTG CCT
ATA AAC CAA CGG TTT ATT TTA GTT GAA CTC TCA AAT CAT GCA GAG CTT TCT
GTT ACA TTA GCC CTG GAT GTC ACC AAT GCA TAT GTG GTC GGC TAC CGT GCT
GGA AAT AGC GCA TAT TTC TTT CAT CCT GAC AAT CAG GAA GAT GCA GAA GCA
ATC ACT CAT CTT TTC ACT GAT GTT CAA AAT CGA TAT ACA TTC GCC TTT GGT
GGT AAT TAT GAT AGA CTT GAA CAA CTT GCT GGT AAT CTG AGA GAA AAT ATC
GAG TTG GGA AAT GGT CCA CTA GAG GAG GCT ATC TCA GCG CTT TAT TAT TAC
AGT ACT GGT GGC ACT CAG CTT CCA ACT CTG GCT CGT TCC TTT ATA ATT TGC
ATC CAA ATG ATT TCA GAA GCA GCA AGA TTC CAA TAT ATT GAG GGA GAA ATG
CGC ACG AGA ATT AGG TAC AAC CGG AGA TCT GCA CCA GAT CCT AGC GTA ATT
ACA CTT GAG AAT AGT TGG GGG AGA CTT TCC ACT GCA ATT CAA GAG TCT AAC
CAA GGA GCC TTT GCT AGT CCA ATT CAA CTG CAA AGA CGT AAT GGT TCC AAA
TTC AGT GTG TAC GAT GTG AGT ATA TTA ATC CCT ATC ATA GCT CTC ATG GTG
TAT AGA TGC GCA CCT CCA CCA TCG TCA CAG TTT

GCT GAT GTT TGT ATG GAT CCT GAG CCC ATA GTG

CGT ATC GTA GGT CGA AAT GGT CTA TGT GTT GAT GTT AGG GAT GGA AGA TTC
CAC AAC GGA AAC GCA ATA CAG TTG TGG CCA TGC AAG TCT AAT ACA GAT GCA
AAT CAG CTC TGG ACT TTG AAA AGA GAC AAT ACT ATT CGA TCT AAT GGA AAG
TGT TTA ACT ACT TAC GGG TAC AGT CCG GGA GTC TAT GTG ATG ATC TAT GAT
TGC AAT ACT GCT GCA ACT GAT GCC ACC CGC TGG CAA ATA TGG GAT
AAT GGA ACC ATC ATA AAT CCC AGA TCT AGT CTA GTT TTA GCA GCG ACA TCA

EP 0 145 111 B1

GGG AAC AGT GGT ACC ACA CTT ACG GTG CAA ACC AAC ATT TAT GCC GTT AGT

CAA GGT TGG CIT CCT ACT AAT AAT ACA CAA CCT TTT GTT ACA ACC ATT GTT

GGG CTA TAT GGT CTG TGC TTG CAA GCA AAT AGT GGA CAA GTA TGG ATA GAG

GAC TGT AGC AGT GAA AAG GCT GAA CAA CAG TGG GCT CTT TAT GCA GAT GGT

TCA ATA CGT CCT CAG CAA AAC CGA GAT AAT TGC CTT ACA AGT GAT TCT AAT

ATA CGG GAA ACA GTT GTT AAG ATC CTC TCT TGT GGC CCT GCA TCC TCT GGC

CAA CGA TGG ATG TTC AAG AAT GAT GGA ACC ATT TTA AAT TTG TAT AGT GGA

TTG GTG TTA GAT GTG AGG CGA TCG GAT CCG AGC CTT AAA CAA ATC ATT CTT

TAC CCT CTC CAT GGT GAC CCA AAC CAA ATA TGG TTA CCA TTA TTT

or at least a portion of a nucleotide sequence which is equivalent thereto by virtue of degeneracy of the genetic code.

30 According to another aspect of this invention we provide a recombinant DNA molecule comprising a DNA sequence which codes for a polypeptide sequence present in a plant toxin of the ricin type, as defined herein.

More specifically, we provide a recombinant DNA molecule containing a DNA sequence which codes for a A and B chain precursor polypeptide of a plant toxin of the ricin type.

35 Alternatively, we provide a recombinant DNA molecule containing a DNA sequence coding for at least a portion of the A chain of a plant toxin of the ricin type said portion coding for a polypeptide that displays the ribosome-inactivity property of the A chain of ricin.

According to a further aspect of the present invention we provide a genetically modified host micro-organism which contains a recombinant DNA molecule as hereinbefore defined.

40 In the recombinant DNA molecules according to the present invention, the nucleotide sequence coding for the B chain may be modified to eliminate or inactivate the galactose binding sites, and the precursor polypeptide and hence in the mature protein, to eliminate or inactivate the galactose binding sites, and the sequences encoding the signals for N-glycosylation may also be modified to render them ineffective or to eliminate them. Examples of techniques which may prove useful are deletion or oligonucleotide mediated

45 mutagenesis. The host organism may be a plant cell or an animal cell or preferably a microorganism.

The microorganism may be a prokaryote or a eukaryote. As examples of prokaryotes may be mentioned Gram-negative bacteria, e.g. *E. coli*, *Methylophilus*, *methylophilus* and *Alcaligenes eutrophus*; and Gram-positive bacteria, e.g. *Streptomyces*, *Bacillus subtilis* and *Arthrobacter*. As examples

50 of eukaryotes may be mentioned yeasts, for example *Saccharomyces cerevisiae*. The recombinant DNA molecules may comprise a cloning vector such as a plasmid or phage vector into which has been inserted the DNA sequence coding for at least a portion of a precursor polypeptide, or at least a portion of either the A chain or the B chain, of a ricin-type plant toxin.

The cloning vector is preferably a plasmid although we do not exclude the possibility that it may be a

55 phage vector. The plasmid may be a naturally-occurring plasmid or preferably a composite derived from fragments of other plasmids. Where a composite plasmid is employed it preferably contains promoter sequences which improve expression of the ricin gene.

Examples of suitable plasmids which may be used as cloning vehicles are *inter alia* for Gram-negative bacteria: pBR322, pAT153, pUC8, pGSS15 and pMB9; for Gram-positive bacteria: pVC6; and for

60 *S.cerevisiae*: pMA91, pMA230, YRp7, pLC544 and YEp6. The vector will be selected to be suitable for the particular host envisaged.

We also provide a method of obtaining a recombinant DNA molecule which comprises preparing a double-stranded DNA sequence which codes for a polypeptid sequence present in a plant toxin of the ricin type and inserting said double-stranded DNA sequence into a cloning vector.

65 More specifically such a method may comprise isolating the mRNA which codes for the ricin A and B

chain precursor polypeptide, synthesising single-stranded cDNA from said mRNA using reverse transcriptase and a suitable primer, assembling the second DNA strand onto the template formed by said first strand by means of DNA polymerase followed by S1 nuclease, and inserting the resulting double-stranded cDNA into a cloning vector.

- 5 Alternatively, the cDNA assembled from the mRNA may be cut into separate portions which code respectively for separate portions of the ricin molecule, for example for the A and B chains, which portions are then inserted into separate cloning vectors.

As stated above the cloning vector is preferably a plasmid such as pBR322, pAT153, of pUC8, and this may be cut open by means of the restriction endonuclease *Pst* I, and may be tailed with oligo (dG), and
10 annealed with the double-stranded cDNA which has been tailed with oligo (dC).

We also provide a method of producing a modified transformed host by introducing into a suitable host microorganism a recombinant DNA molecule according to this invention.

The microorganism used as the host for cloning is preferably a Gram-negative bacterium and more preferably *E. coli*.

- 15 After cloning, the DNA sequence coding for the ricin precursor (or the precursor of another ricin-type toxin which is formed from a precursor) may be removed from the host cloning vector. It may then be divided into two portions which code for separate regions of the toxin molecule, for example the A and B chains, these portions introduced into separate second cloning vectors and new hosts modified with each of the resulting new recombinant DNA molecules. Alternatively it may be introduced whole into a second
20 cloning vector. The second cloning vectors comprise suitable promoter sequences and the position and direction of insertion of the whole coding sequence or portion thereof into the second cloning vector are such that on introducing the new recombinant DNA molecules into suitable host microorganisms, e.g. *E. coli* or *S. cerevisiae*, expression of the desired gene sequence is obtained.

The preparation of a transformed host containing a DNA sequence which codes for the ricin A and B
25 chain precursor polypeptide will now be described by way of example, first in general terms and then in detail. This process is summarised in the accompanying diagram.

- Firstly, the mRNA encoding this precursor was enriched in known manner by sucrose density gradient centrifugation. The corresponding cDNA was assembled in single-strand form onto this mRNA in known manner using the enzyme reverse transcriptase, a growing point having first been provided on the mRNA
30 using oligo (dT) as primer which binds onto the polyadenylated 3'-terminus of the mRNA. The immediate product of this reaction is a DNA-RNA hybrid. The RNA strand is removed by hydrolysis, leaving the single-strand DNA intact. This is converted to the double stranded form using the enzyme DNA polymerase in the presence of free nucleotides, which results in a hairpin-shaped molecule, the curved end of which is then removed by the single-strand specific nuclease S1. The resulting double strand cDNA is then tailed
35 with oligo (dC) using terminal transferase, size fractionated to remove small molecules, or vice versa and annealed with the pBR322 or pAT153 vector which has been cut open with *Pst* I and tailed with oligo (dG) using terminal transferase, the cytosine tails on the DNA base pairing with the guanine tails on the vector.

- The resulting chimaeric plasmids containing the DNA segment coding for the ricin precursor polypeptide were then used to transform *E. coli* DH1 cells, and the presence of the chimaeric plasmid was
40 ensured by selecting cells displaying tetracycline resistance and ampicillin sensitivity. Over 1600 Tet^r, Amp^r clones were obtained. Colonies derived from each clone were transferred to nitrocellulose filters and clones containing the desired DNA sequence identified using a 32P-end labelled 20 mer oligonucleotide probe, having the DNA sequence ACCTACAAATTCTCTCC which hybridises to DNA containing the complementary sequence TGGATGTTAAAGAGG. As the ricin precursor polypeptide have been
45 found to contain the amino sequence — Trp-Met-Phe-Lys-Asn-Asp-Gly- DNA sequence responsible for this is known from the genetic code to be the latter mentioned above.

- Using appropriate hybridisation and wash conditions, e.g. as described by Singer-Sam *et al* in (1983) Proc. Natl. Acad. Sci. (U.S.A.), Vol. 80 pp 802—806, 80 clones were selected as positively containing the
50 desired DNA sequence, and of these, the eight largest in the plasmid pBR322 have been initially chosen for further characterisation. Their relationship to the castor bean lectin precursor polypeptides has been confirmed using the hybrid release translation assay. Of the eight clones mentioned above, four respectively with 1614, 1950, 1059 and 1020 base pairs, have been selected for sequencing.

In detail, the transformed host was prepared as follows:

55 A. cDNA synthesis

1. mRNA extraction and fractionation

- 100—200 g of ripening *Ricinus* seeds were frozen and ground to a powder in liquid nitrogen, and homogenised in a Waring blender for 1—2 minutes in 50 mM tris-HCl pH 9, 150 mM NaCl, 5 mM EDTA and 5% SDS. The homogenate was extracted with an equal volume of phenol:chloroform (1:1) and the phases
60 were separated by centrifugation. The organic phase and residue were reextracted with 0.5 volume of 20 mM tris-HCl pH 9.0, 2 mM EDTA and the resultant aqueous phase was combined with the original one. The total aqueous phase was reextracted repeatedly with equal volumes of phenol:chloroform until no material was present at the interface. RNA was precipitated by the addition of 2 volumes of cold ethanol after making the solution 200 mM in NaCl.

65

After overnight precipitation at -20°C the RNA was centrifuged at 10,000 rpm for 30 minutes in an MSE 18 or MSE 21 centrifuge; the pellet was then washed repeatedly in 3 M NaAc pH 5.5 until no polysaccharide was detectable in the supernatant by ethanol precipitation. The final pellet was dissolved in 300 mM NaCl and precipitated as above.

5 mRNA molecules bearing poly(A) tails were extracted by affinity chromatography on olig (dT)-cellulose: after hybridisation at room temperature for 30 min. in 400 mM NaCl, 20 mM tris-HCl pH 7.6, 0.2% SDS, the beads were pelleted and washed three times in the above buffer and two times in 200 mM NaCl, 20 mM tris-HCl pH 7.6, 0.1% SDS. The slurry was poured into a column and washed further with the last buffer until the A_{260} of the eluate reached the background level. Poly(A)-containing RNA was then eluted
10 with 20 mM tris-HCl pH 7.6 at 50°C . The eluate was monitored with an ISCO continuous flow UV cell. Poly(A)-containing RNA was precipitated overnight from 200 mM NaCl by the addition of 2 volumes of cold ethanol at -20°C and was then washed three times with 70% ethanol, and redissolved in 10 mM tris-HCl pH 7.0 to approximately 1 $\mu\text{g}/\mu\text{l}$.

The mRNA was heated for 2 min. at 65°C and quench cooled. Approximately 400 μg of poly(A)⁺ RNA
15 was layered on top of a 10—30% ribonuclease-free sucrose (Sigma) density gradient in 100 mM tris-HCl pH 7.5, 0.5% SDS, 1 mM EDTA, and centrifuged in a Beckman L5—65B centrifuge using an SW27 rotor at 25,000 rpm at 17°C for 14 hours. 400 μl fractions were collected with an ISCO density gradient fractionator using the continuous flow UV cell.

Each fraction was made 200 mM in NaCl and precipitated with 2 volumes of cold ethanol by freeze-
20 thawing in liquid nitrogen three times, and recovered by centrifugation in an Eppendorf microfuge for 30 minutes at 4°C , washed once with 70% ethanol, and redissolved in 10 μl of 10 mM tris-HCl pH 7.0. An aliquot from each fraction (1 μl) was translated in a reticulocyte lysate cell free system and the lectin precursor immunoprecipitated to identify the fraction enriched for lectin mRNA.

25 2. First strand synthesis

Fractionated poly(A)⁺ RNA was reverse transcribed at 0.5 $\mu\text{g}/\mu\text{l}$ in the presence of 50 mM tris-HCl pH 8.3, 10 mM MgCl_2 , 100 mM KCl, 1 mM of dATP, dTTP and dGTP, 250 μM dCTP, 0.06 $\mu\text{g}/\mu\text{l}$ oligo(dT)₁₂₋₁₈, 10 mM DTT and 0.4 units/ μl of reverse transcriptase from avian myeloblastosis virus. (3H)dCTP or α -^{(32)P}dCTP were included in the reaction as appropriate.

30 The reaction mixture was incubated at 42°C for 45 minutes, at which point an equal volume of 5 mM tris-HCl pH 8.3, 5 mM DTT, 250 μM dCTP was added along with the same amount of enzyme as previously. The reaction was incubated for a further 45 minutes at 45°C and terminated by freezing. Aliquots were analysed on 1% denaturing agarose gels along with the products of the second strand and S_1 nuclease reactions.

35

3. Second strand synthesis

mRNA—cDNA hybrids were denatured by boiling the first strand reaction for 3 minutes and cooling rapidly. After pelleting insoluble material in the Eppendorf microfuge for 2 minutes the supernatant was transferred to a fresh chilled tube. For the standard reaction, reagents were added as follows, ignoring
40 elements already present: dATP, dGTP and dTTP to 100 μM , Hepes-KOH pH 6.9 to 105 mM, KCl to 92 mM, dCTP, labelled as appropriate, to 80 μM , and 0.1 units/ μl of DNA polymerase. The reaction was allowed to proceed at 20°C for 6 hours, at which time cDNA was removed from the mixture by gel filtration on 1 ml columns of Bio-Gel P60 in 10 mM tris-HCl pH 7.6, 20 mM NaCl, 1 mM EDTA. Fractions were monitored by Cerenkov or liquid scintillation counting, and peak excluded fractions were pooled and precipitated from
45 0.3 M NaAc pH 6 by the addition of 2 volumes of cold ethanol. Precipitates were recovered by centrifugation in the Eppendorf microfuge for 30 minutes in the cold, and dissolved in water to about 2.5 $\mu\text{g}/\mu\text{l}$ of RNA — equivalent material.

4. S_1 nuclease digestion

50 Single-stranded regions of double-stranded cDNA were digested with S_1 nuclease from *Aspergillus oryzae*, in the presence of 300 mM NaCl, 30 mM NaAc pH 4.5, 3 mM ZnCl_2 . The reaction was incubated for 15 minutes at 37°C and then for 15 minutes at 15°C , and was terminated by the addition of tris-HCl pH 7.6 to 130 mM and EDTA to 10 mM; it was then extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated from 300 mM NaAc pH 6 with 2 volumes of cold ethanol. The precipitate
55 was dissolved in 10 mM tris-HCl pH 8, 0.1 mM EDTA to 0.25 $\mu\text{g}/\mu\text{l}$ RNA equivalent.

5. Addition of homopolymer tails to DNA

Double-stranded DNA was tailed using terminal transferase with dCTP at 0.001—0.01 $\mu\text{g}/\mu\text{l}$ in the presence of 140 mM potassium cacodylate pH 7.6, 30 mM tris base, 0.1 mM DTT, 1 mM CoCl_2 and (3H) or
60 (³²P) — labelled dCTP in 75—150 fold excess over 3' termini. The reaction was carried out at 37°C for 6 minutes. The extent of incorporation of label was followed by assaying the amount of TCA — insoluble radioactivity as a proportion of the total radioactivity, counting in Bray's scintillant.

The reaction was stopped by chilling and adding EDTA to 10 mM, after which unincorporated material was removed by gel filtration as described. Tailed cDNA was precipitated as before, and dissolved in 1 M
65 NaAc pH 8, 10 mM tris-acetate pH 8, 1 mM EDTA ready for fractionation.

PstI-cleaved pBR322 DNA was similarly treated, except that dGTP replaced dCTP.

6. Fractionation of tailed cDNA

cDNA was fractionated on 5–20% linear sucrose density gradients in 1 M NaAc pH 8, 10 mM tris-acetate pH 8, 1 mM EDTA, and centrifuged overnight at 39,000 rpm in an SW50.1 rotor. DNA sedimentation was checked on a parallel gradient loaded with a mixture of HinfI and PstI digests of pBR322 DNA, and fractions of this gradient were run on a 1% neutral agarose gel. Fractions from the cDNA gradient were diluted with an equal volume of water and precipitated with 2 volumes of cold ethanol, and then pooled to give three final fractions, a large cDNA fraction (larger than 2,200 bp), an intermediate fraction (1,000–2,200 bp) and a fraction containing smaller cDNAs (600–1,000 bp). cDNA molecules smaller than 600 bp were discarded.

The three final fractions were dissolved to approximately 5 ng/μl in 150 mM RbCl, 10 mM tris-HCl pH 7.6, 0.2 mM EDTA.

B. Annealing and transformation

1. Annealing

dC-tailed cDNA was mixed with dG-tailed pBR322 or pAT153 in approximately equimolar quantities, at a concentration of 0.4 ng/μl of vector. Buffers were as described above. The mixtures were heated to 70°C for 30 minutes and then cooled overnight to room temperature, and slowly chilled to 4°C. Competent cells were added and transformed as described below.

2. Preparation of competent cells and transformation

DHI cells (recA⁺, nalA⁺, r⁺_k, m⁺_k, endoI⁺, R⁺, relA17) were grown in 10 ml cultures of psi broth (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 20 mM MgCl₂, pH adjusted to 7.6 with KOH; all bacteriological reagents from Difco), and grown at 37°C in a shaking waterbath to A₅₅₀ = 0.3. 1 ml of this was then inoculated into 25 ml of the same medium and grown to A₅₅₀ = 0.48. The cells were then chilled on ice for 15 minutes and harvested at 5,000 rpm for 5 minutes in an MSE 21 centrifuge at 4°C. They were then resuspended in 10 ml of 100 mM RbCl, 50 mM MgCl₂, 10 mM CaCl₂, 35 mM NaAc pH 5.8, 15% glycerol and kept on ice for 15 minutes.

The cells were again harvested, and resuspended in 1 ml of 10 mM RbCl, 75 mM CaCl₂, 10 mM MOPS-KOH pH 5.8, 15% glycerol, and kept on ice for a further 15 minutes.

100 μl of the cells thus prepared were mixed with the annealed DNA samples, and incubated on ice for 30 minutes, after which they were heat-shocked at 42°C for 90–120 seconds. 1 ml of psi broth was added, and the cells were grown at 37°C for 1 hour. They were then centrifuged briefly, and resuspended in 100 μl of psi broth and plate on LB plates containing 14 μg/ml tetracycline (LB is 1% tryptone, 0.5% yeast extract, 170 mM NaCl, 1.5% agar).

After 18–24 hours growth at 37°C, colonies were counted and spotted onto LB plates containing 33 μg/ml ampicillin to identify those transformants containing recircularised or uncut plasmids. Over 1600 Tet^r Amp^r clones were picked and transferred in ordered arrays onto large plates of LB containing 14 μg/ml tetracycline.

C. Screening

1. Labelling of oligonucleotide

The ricin B chain specific oligomer (20 mer) was end labelled using polynucleotide kinase. 500 ng of oligonucleotide was incubated in 50 mM tris pH 8.5, 10 mM MgCl₂, 5 mM DTT, 0.1 mM spermidine-HCl, 0.1 mM EDTA with 60 μCi γ(³²P) ATP and 1 μl polynucleotide kinase (Boehringer) for 35 min at 37°C. The reaction was stopped by adding an equal volume of 0.6 M NH₄AC and the bulk of non incorporated γATP was removed by passage through a sephadex G25 column in 0.14 M NaCl, 0.02 M tris pH 7.6, 0.005 M EDTA, 0.1% SDS. The probe was stored frozen at –20°C.

2. Colony Hybridization using oligonucleotide probe

Transformants were grown on nitrocellulose filters (Schleicher & Schuell 0.45 μ) layered over LB plus tetracycline. The filters, in triplicate, were then transferred to LB-Tet plates containing 200 μg/ml chloramphenicol for 16 h at 37°. The filters were placed colony side up on two sheets of 3 mm paper wetted with 0.5 M NaOH for 15 mins at room temperature. The same procedure was followed for the following two washes (1) with 1 M tris pH 8.0, and (2) with 1 M tris pH 8, 1.5 M NaCl (30 mins). The filters were air dried and baked at 80°C.

Prehybridizations and hybridizations were done in double sealed polythene bags. The filters were prehybridized in 0.9 M NaCl, 0.09 M tris 7.4, 0.006 M EDTA, 0.5% NP40, 2 × Denhardt's, 0.2% SDS, 100 μg/ml denatured single strand salmon sperm DNA and 70 μg/ml tRNA. Prehybridization was done for 4 h at 55°C. The prehybridization buffer was then squeezed from the bag and fresh buffer added that contained 50 ng labelled probe (to give a maximum concentration of 5 ng/ml buffer). Annealing was done overnight at 37°C.

Non stringent washes were done with 6 × SSC at room temperature. The filters were washed in 4 changes of 6 × SSC over 3 h. The triplicate filters were then washed at three different temperatures determined from the base composition and degree of mismatch of the probe. Using 2°C for every A or T

and 4°C for every C or G in the probe the wash temperatures selected were 52°C, 56°C and 60°C. The filters were washed at the stringent temperature in $6 \times$ SSC for 10 minutes and then dried thoroughly. The filters were exposed to X-ray film overnight.

5 D. Hybrid selection procedure

1. DNA binding

Plasmid DNA was purified from the positive clone(s) and 10–15 µg linearized with EcoRI. After phenol:chloroform extraction and ethanol precipitation the pellet was dissolved in 0.5 ml $0.1 \times$ SSC. 0.5 ml 1 M NaOH was then added and the mixture allowed to stand for 15 mins at room temperature. 4 ml of a
10 prechilled neutralizing solution (1.5 M NaCl, 0.25 M HCl, 0.25 M tris-(pH 8.0)) was added and the 5 ml DNA sample sucked by vacuum through swinnies containing wetted Schleicher and Schuell 0.45 µ filter discs. 5 ml $6 \times$ SSC was then passed through the filter(s). These were air dried and then baked for 2 h at 80°C.

2. Hybrid selection protocol

15 The filter(s) were placed in 5 ml bottles and prehybridized for 4 h at 41°C in 50% formamide, 0.4 M NaCl, 10 mM pipes-NaOH pH 6.4, 4 mM EDTA, 0.5 µg/ml tRNA, 10 µg/ml poly (A). The buffer was removed and the filter(s) typically hybridized overnight at 41°C in 50% formamide buffer (above) containing approx. 20 µg poly (A)* RNA from castor beans. The buffer was removed and the filters washed twice for 15 mins in each of the following: (1) $1 \times$ SSC, 0.5% SDS at room temperature, (2) $0.1 \times$ SSC, 0.1% SDS at room
20 temperature, (3) $0.1 \times$ SSC, 0.1% SDS at 50°C, (4) $0.1 \times$ SSC, 0.1% SDS at room temperature. The filters were drained and 200 µl hybrid release buffer (90% formamide, 10 mM pipes-NaOH pH 6.4, 1 mM EDTA, 0.5% SDS) added to each and mixed for 30 mins at 40°C. The buffer was removed into a fresh eppendorf and NaCl added to .2 M. The released mRNA was precipitated with ethanol, rinsed several times in 70% ethanol, dried and dissolved in 5 µl sterile water. The sample was translated in a reticulocyte lysate cell free
25 system and the products run out directly on an SDS-polyacrylamide gel or firstly immunoprecipitated with appropriate antisera.

The DNA sequence coding for the above-mentioned ricin precursor polypeptide of two of the above-mentioned clones, referred to herein as pBRCL 6 and pBRCL 17 (RCL = Ricinus communis lectin) has now been determined by a combination of the Sanger dideoxy method (Sanger et al, 1977 — Proc. Natl. Acad. Sci. U.S.A. 74, 5463–67) and the procedure of Maxam and Gilbert (Maxam and Gilbert, 1980 — Meth. Enzym. 65, 499–560). In order to determine the sequence at the ends of each insert, the inserts were excised from pBR322 with Pst I and ligated into Pst I linearized, phosphatased plasmid pUC8 (Vierra and
30 Messing, 1982 — Gene 19, 259–268). E. Coli DH1 cells were transformed by these recombinant plasmids. These new recombinant plasmids are referred to herein as pRCL6 and pRCL17.

35 It is apparent that the two inserts contain a region of common sequence and that together they represent a total ricin precursor sequence. There are no nucleotide differences between the overlapping regions of the inserts in pRCL6 and pRCL17.

A new recombinant DNA molecule was then constructed that contains the complete nucleotide sequence encoding the ricin precursor polypeptide. This was achieved by isolating a fragment, 323 base
40 pairs in length, obtained from pRCL17 by digestion with the restriction endonuclease Sau 961, and ligating this fragment to a fragment 1561 base pairs in length isolated after a partial digestion of pRCL6 with Sau 961. Ligation was performed in 50 mM tris HCl (pH 7.4) 10 mM MgCl₂, 10 mM dithiothreitol, 1 mM spermidine, 10 mM ATP, 0.1 mg/ml BSA with 5 units commercial T4 DNA ligase, and incubation proceeded overnight at 15°C. After a standard phenol/chloroform extraction and ethanol precipitation the ligated DNA
45 was pelleted, dissolved in a small volume of 10 mM tris HCl (pH 7.4), 1 mM EDTA and digested to completion with Pst I. The resulting linearised DNA was then ligated (as above) with an equal quantity of Pst I linearised, phosphatased pUC8. The new recombinant DNA molecule containing the entire DNA sequence of the ricin precursor and referred to as pRCL617, was used in conventional manner to transform E. coli DH1 cells.

50 The nucleotide sequence of pRCL617 is shown hereinafter.

This sequence was deduced from the two overlapping cDNA inserts in clones pRCL6 and pRCL17 (the limits of the DNA inserts in each of these two clones are given below).

Nucleotide residues are numbered in the 5' to 3' direction with the first residue of the codon specifying the amino terminal residue of mature ricin A chain numbered 1 and the nucleotides on the 5' side of residue
55 1 indicated by negative numbers. The 5' terminal sequence does not extend to the 5' end of the mRNA whereas the 3' terminal sequence shown is followed by a poly (dA) tract 27 residues long, thus representing the complete sequence of the region. The predicted amino acid sequence is given below the nucleotide sequence and differences with the published amino acid sequence of mature ricin A and B chains (Funatsu G., Kimura, M. and Funatsu, M. Agric. Biol. Chem. Vol 43, pp 2221–2224 (1979), and
60 Yoshitake, S., Funatsu, G. and Funatsu, M. — Agric. Biol. Chem. Vol. 42, pp 1267–1274 (1978)) are indicated und meath. Residues absent from the published amino acid sequence are underlined with a dashed line and the position of amino acids present in the published sequence but absent from the derived sequence pres nted here are indicated by an asterisk. The dashed line beneath the 12 amino acid sequence linking the C-terminus of the A chain and the N-terminus of the B chain is bracketed. Amino acids are numbered
65 from the amino terminal residue of th mature A chain and the preceding residu s are indicated by

EP 0 145 111 B1

negative numbers. Potential sites for asparagine linked N-glycosylation are boxed and potential poly (A) signals are underlined. The insert of pRCL6 extends from nucleotide — 102 to residue 1512 and the insert of pRCL17 extends from nucleotide 733 to residue 1782.

The intervening twelve triplets code for the linker amino acid sequence which is present in the precursor polypeptide and which is enzymatically removed in the cell to separate the A and B chains, which chains are joined by a disulphide bridge during the formation of the ricin molecule itself. This linker region as well as the presumptive amino terminal leader or signal sequence (amino acids —24 to —1) are not present in the sequences already published by Funatsu *et al.*

Preproricin is the whole polypeptide coded for the aforesaid DNA insert, i.e. from amino acid —24 to amino acid 541. Proricin, which is obtained from preproricin in the organism by removal of the amino acid leader sequence, extends from amino acid 1 to amino acid 541.

```

-100                                     -50
5'-AAACCGGGAG GAATACTAT TGTAATATGG ATG TAT GCA GTG GCA ACA TGG CTT
15                               Met Tyr Ala Val Ala Thr Trp Leu
                                   -20

                                   -1 1
20 TGT TTT GGA TCC ACC TCA GGG TGG TCT TTC ACA TTA GAG GAT AAC AAC ATA
   Cys Phe Gly Ser Thr Ser Gly Trp Ser Phe Thr Leu Glu Asp Asn Asn Ile
                                   -1 1
                                   -10

25 TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGT GCC ACT GTG
   Phe Pro Lys Gln Tyr Pro Ile Ile Asn Phe Thr Thr Ala Gly Ala Thr Val
                                   50
                                   10

30 CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TTA ACA ACT GGA
   Gln Ser Tyr Thr Asn Phe Ile Arg Ala Val Arg Gly Arg Leu Thr Thr Gly
                                   20
                                   30

35 GCT GAT GTG AGA CAT GAT ATA CCA GTG TTG CCA AAC AGA GTT GGT TTG CCT
   Ala Asp Val Arg His Asp Ile Pro Val Leu Pro Asn Arg Val Gly Leu Pro
                                   40
                                   50

                                   150

40 GCT GAT GTG AGA CAT GAT ATA CCA GTG TTG CCA AAC AGA GTT GGT TTG CCT
   Ala Asp Val Arg His Asp Ile Pro Val Leu Pro Asn Arg Val Gly Leu Pro
                                   40
                                   50

                                   200

45 ATA AAC CAA CGG TTT ATT TTA GTT GAA CTC TCA AAT CAT GCA GAG CTT TCT
   Ile Asn Gln Arg Phe Ile Leu Val Glu Leu Ser Asn His Ala Glu Leu Ser
                                   60
                                   Gln

                                   250

50 GTT ACA TTA GCC CTG GAT GTC ACC AAT GCA TAT GTG GTC GGC TAC CGT GCT
   Val Thr Leu Ala Leu Asp Val Thr Asn Ala Tyr Val Val Gly Tyr Arg Ala
   70                               Ser                               80

                                   300

55 GGA AAT AGC GCA TAT TTC TTT CAT CCT GAC AAT CAG GAA GAT GCA GAA GCA
   Gly Asn Ser Ala Tyr Phe Phe His Pro Asp Asn Gln Glu Asp Ala Glu Ala
   90                               100

                                   350

60 ATC ACT CAT CTT TTC ACT GAT GTT CAA AAT CGA TAT ACA TTC GCC TTT GGT
   Ile Thr His Leu Phe Thr Asp Val Gln Asn Arg Tyr Thr Phe Ala Phe Gly
   110                               120

```

EP 0 145 111 B1

400
 GGT AAT TAT GAT AGA CTT GAA CAA CTT GCT GGT AAT CTG AGA GAA AAT ATC
 Gly Asn Tyr Asp Arg Leu Glu Gln Leu Ala Gly Asn Leu Arg Glu Asn Ile
 130

450
 GAG TTG GGA AAT GGT CCA CTA GAG GAG GCT ATC TCA GCG CTT TAT TAT TAC
 Glu Leu Gly Asn Gly Pro Leu Glu Glu Ala Ile Ser Ala Leu Tyr Tyr Tyr
 140 150

500
 AGT ACT GGT GGC ACT CAG CTT CCA ACT CTG GCT CGT TCC TTT ATA ATT TGC
 Ser Thr Gly Gly Thr Gln Leu Pro Thr Leu Ala Arg Ser Phe Ile Ile Cys
 160 170

550
 ATC CAA ATG ATT TCA GAA GCA GCA AGA TTC CAA TAT ATT GAG GGA GAA ATG
 Ile Gln Met Ile Ser Glu Ala Ala Phe Gln Tyr Ile Glu Gly Glu Met
 180

600
 CGC ACG AGA ATT AGG TAC AAC CGG AGA TCT GCA CCA GAT CCT AGC GTA ATT
 Arg Thr Arg Ile Arg Tyr Asn Arg Arg Ser Ala Pro Asp Pro Ser Val Ile
 190 200

650
 ACA CTT GAG AAT AGT TGG GGG AGA CTT TCC ACT GCA ATT CAA GAG TCT AAC
 Thr Leu Glu Asn Ser Trp Gly Arg Leu Ser Thr Ala Ile Gln Glu Ser Asn
 210 220

700
 CAA GGA GCC TTT GCT AGT CCA ATT CAA CTG CAA AGA CGT AAT GGT TCC AAA
 Gln Gly Ala Phe Ala Ser Pro Ile Gln Leu Gln Arg Arg Asn Gly Ser Lys
 230 --- Asp

750
 TTC AGT GTG TAC GAT GTG AGT ATA TTA ATC CCT ATC ATA GCT CTC ATG GTG
 Phe Ser Val Tyr Asp Val Ser Ile Leu Ile Pro Ile Ile Ala Leu Met Val
 240 250 ---

800
 TAT AGA TGC GCA CCT CCA CCA TCG TCA CAG TTT TCT TTG CTT ATA AGG CCA
 Tyr Arg Cys Ala Pro Pro Pro Ser Ser Gln Phe Ser Leu Leu Ile Arg Pro
 260 (--- --- --- --- --- ---

850
 GTG GTA CCA AAT TTT AAT GCT GAT GTT TGT ATG GAT CCT GAG CCC ATA GTG
 Val Val Pro Asn Phe Asn Ala Asp Val Cys Met Asp Pro Glu Pro Ile Val
 --- --- --- --- --- ---) 280 290

900
 CGT ATC GTA GGT CGA AAT GGT CTA TGT GTT GAT GTT AGG GAT GGA AGA TTC
 Arg Ile Val Gly Arg Asn Gly Leu Cys Val Asp Val Arg Asp Gly Arg Ph
 300 Asn

EP 0 145 111 B1

950

CAC AAC GGÀ AAC GCA ATA CAG TTG TGG CCA TGC AAG TCT AAT ACA GAT GCA
His Asn Gly Asn Ala Ile Gln Leu Trp Pro Cys Lys Ser Asn Thr Asp Ala
Asn His 310 320

1000

AAT CAG CTC TGG ACT TTG AAA AGA GAC AAT ACT ATT CGA TCT AAT GGÀ AAG
Asn Gln Leu Trp Thr Leu Lys Arg Asp Asn Thr Ile Arg Ser Asn Gly Lys
--- 330 340

1050

TGT TTA ACT ACT TAC GGG TAC AGT CCG GGA GTC TAT GTG ATG ATC TAT GAT
Cys Leu Thr Thr Tyr Gly Tyr Ser Pro Gly Val Tyr Val Met Ile Tyr Asp
Pro Ser

1100

TGC AAT ACT GCT GCA ACT GAT GCC ACC CGC TGG CAÀ ATA TGG GAT
Cys Asn Thr Ala Ala Thr Asp Ala Thr Arg Trp Gln Ile Trp Asp
360 Thr Asp --- Glu Asn

1150

AAT GGA ACC ATC ATA AAT CCC AGA TCT AGT CTA GTT TTA GCA GCG ACA TCA
Asn Gly Thr Ile Ile Asn Pro Arg Ser Ser Leu Val Leu Ala Ala Thr Ser
380 390

1200

GGG AAC AGT GGT ACC ACA CTT ACG GTG CAÀ ACC AAC ATT TAT GCC GTT AGT
Gly Asn Ser Gly Thr Thr Leu Thr Val Gln Thr Asn Ile Tyr Ala Val Ser
400

1250

CAA GGT TGG CTT CCT ACT AAT AAT ACA CAÀ CCT TTT GTT ACA ACC ATT GTT
Gln Gly Trp Leu Pro Thr Asn Asn Thr Gln Pro Phe Val Thr Thr Ile Val
Pro Phe Trp 420

1300

GGG CTA TAT GGT CTG TGC TTG CAA GCA AAT AGT GGA CAÀ GTA TGG ATÀ GAG
Gly Leu Tyr Gly Leu Cys Leu Gln Ala Asn Ser Gly Gln Val Trp Ile Glu
430 Val 440

1350

GAC TGT AGC AGT GAA AAG GCT GAA CAÀ CAG TGG GCT CTT TAT GCA GAT GGT
Asp Cys Ser Ser Glu Lys Ala Glu Gln Gln Trp Ala Leu Tyr Ala Asp Gly
Ser Cys 450 Ser

1400

TCA ATÀ CGT CCT CAG CAÀ AAC CGA GAT AAT TGC CTT ACA AGT GAT TCT AAT
Ser Ile Arg Pro Gln Gln Asn Arg Asp Asn Cys Leu Thr Ser Asp Ser Asn
Asn 460 Asn Arg 470

1450

ATA CGG GAA ACA GTT GTT AAG ATC CTC TCT TGT GGC CCT GCA TCC TCT GGC
Ile Arg Glu Thr Val Val Lys Ile Leu Ser Cys Gly Pr Ala Ser Ser Gly
480 490

EP 0 145 111 B1

1500
CAA CGA TGG ATG TTC AAG AAT GAT GGA ACC ATT TTA AAT TTG TAT AGT GGA
Gln Arg Trp Met Phe Lys Asn Asp Gly Thr Ile Leu Asn Leu Tyr S r Gly
500

1550
TTG GTG TTA GAT GTG AGG CGA TCG GAT CCG AGC CTT AAA CAA ATC ATT CTT
Leu Val Leu Asp Val Arg Arg Ser Asp Pro Ser Leu Lys Gln Ile Ile Leu
510 Ala 520

1600
TAC CCT CTC CAT GGT GAC CCA AAC CAA ATA TGG TTA CCA TTA TTT TGA
Tyr Pro Leu His Gly Asp Pro Asn Gln Ile Trp Leu Pro Leu Phe ***
Trp * --- --- *Leu Pro

1650
TAGACAGATT ACTCTCTTGC AGTGTGTGTG TCCTGCCATG AAAATAGATG GCTTAAATAA

1700
AAAGGACATT GTAAATTTTG TAACTGAAAG GACAGCAAGT TATTGCAGTC CAGTATCTAA

1750
TAAGAGCACA ACTATTGTCT TGTGCATTCT 1780
AAATTT-Poly(A)

Claims

1. A recombinant DNA molecule characterised in that it includes as an insert DNA coding for the A chain of ricin and having the following nucleotide sequence or a portion thereof coding for a polypeptide that displays the ribosome-inactivity property of the A-chain of ricin:

ATA
40 TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGI GCC ACT GTG
CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TIA ACA ACT GGA
45 GCT GAT GTG AGA CAT GAT ATA CCA GTG TIG CCA AAC AGA GTT GGT TTG CCT
50 ATA AAC CAA CGG TTT ATT TTA GTT GAA CTC TCA AAT CAT GCA GAG CTT TCT
GTT ACA TTA GCC CTG GAT GTC ACC AAT GCA TAT GTG GTC GGC TAC CGT GCT
55 GGA AAT AGC GCA TAT TTC TTT CAT CCT GAC AAT CAG GAA GAT GCA GAA GCA
ATC ACT CAT CTT TTC ACT GAT GTT CAA AAT CGA TAT ACA TTC GCC TTT GGT
60 GGT AAT TAT GAT AGA CTT GAA CAA CTT GCT GGT AAT CTG AGA GAA AAT ATC
65 GAG TTG GGA AAT GGT CCA CTA GAG GAG GCT ATC TCA GCG CTT TAT TAT TAC

EP 0 145 111 B1

AGT ACT GGT GGC ACT CAG CTT CCA ACT CTG GCT CGT TCC TTT ATA ATT TGC

5 ATC CAA ATG ATT TCA GAA GCA GCA AGA TTC CAA TAT ATT GAG GGA GAA ATG

CGC ACG AGA ATT AGG TAC AAC CGG AGA TCT GCA CCA GAT CCT AGC GTA ATT

10 ACA CTT GAG AAT AGT TGG GGG AGA CTT TCC ACT GCA ATT CAA GAG TCT AAC

CAA GGA GCC TTT GCT AGT CCA ATT CAA CTG CAA AGA CGT AAT GGT TCC AAA

15 TTC AGT GTG TAC GAT GTG AGT ATA TTA ATC CCT ATC ATA GCT CTC ATG GTG

20 TAT AGA TGC GCA CCT CCA CCA TCG TCA CAG TTT

or a nucleotide sequence which is equivalent to said sequence or portion by virtue of the degeneracy of the genetic code.

2. A recombinant DNA molecule according to Claim 1 characterised in that it includes as an insert DNA coding for the ricin and having the following nucleotide sequence or a portion thereof coding for a protein
25 that displays the cell binding and ribosome-inactivity property of the ricin:

ATA

30 TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGI GCC ACT GTG

CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TTA ACA ACT GGA

35 GCT GAT GTG AGA CAT GAT ATA CCA GTG TIG CCA AAC AGA GTT GGT TTG CCT

40 ATA AAC CAA CGG TTT ATT TTA GTT GAA CTC TCA AAT CAT GCA GAG CTT TCT

GTT ACA TTA GCC CTG GAT GTC ACC AAT GCA TAT GTG GTC GGC TAC CGT GCT

45 GGA AAT AGC GCA TAT TTC TTT CAT CCT GAC AAT CAG GAA GAT GCA GAA GCA

ATC ACT CAT CTT TTC ACT GAT GTT CAA AAT CGA TAT ACA TTC GCC TTT GGT

50 GGT AAT TAT GAT AGA CTT GAA CAA CTT GCT GGT AAT CTG AGA GAA AAT ATC

GAG TTG GGA AAT GGT CCA CTA GAG GAG GCT ATC TCA GCG CTT TAT TAT TAC

55 AGT ACT GGT GGC ACT CAG CTT CCA ACT CTG GCT CGT TCC TTT ATA ATT TGC

60 ATC CAA ATG ATT TCA GAA GCA GCA AGA TTC CAA TAT ATT GAG GGA GAA ATG

CGC ACG AGA ATT AGG TAC AAC CGG AGA TCT GCA CCA GAT CCT AGC GTA ATT

65 ACA CTT GAG AAT AGT TGG GGG AGA CTT TCC ACT GCA ATT CAA GAG TCT AAC

EP 0 145 111 B1

CAA GGA GCC TTT GCT AGT CCA ATT CAA CTG CAA AGA CGT AAT GGT TCC AAA
5 TTC AGT GTG TAC GAT GTG AGT ATA TTA ATC CCT ATC ATA GCT CTC ATG GTG
TAT AGA TGC GCA CCT CCA CCA TCG TCA CAG TTT TCT TTG CTT ATA AGG CCA
10 GTG GTA CCA AAT TTT AAT GCT GAT GTT TGT ATG GAT CCT GAG CCC ATA GTG
CGT ATC GTA GGT CGA AAT GGT CTA TGT GTT GAT GTT AGG GAT GGA AGA TTC
15 CAC AAC GGA AAC GCA ATA CAG TTG TGG CCA TGC AAG TCT AAT ACA GAT GCA
AAT CAG CTC TGG ACT TTG AAA AGA GAC AAT ACT ATT CGA TCT AAT GGA AAG
20 TGT TTA ACT ACT TAC GGG TAC AGT CCG GGA GTC TAT GTG ATG ATC TAT GAT
TGC AAT ACT GCT GCA ACT GAT GCC ACC CGC TGG CAA ATA TGG GAT
25 AAT GGA ACC ATC ATA AAT CCC AGA TCT AGT CTA GTT TTA GCA GCG ACA TCA
AAT GGA ACC ATC ATA AAT CCC AGA TCT AGT CTA GTT TTA GCA GCG ACA TCA
30 GGG AAC AGT GGT ACC ACA CTT ACG GTG CAA ACC AAC ATT TAT GCC GTT AGT
CAA GGT TGG CIT CCT ACT AAT AAT ACA CAA CCT TTT GTT ACA ACC ATT GTT
35 GGG CTA TAT GGT CTG TGC TTG CAA GCA AAT AGT GGA CAA GTA TGG ATA GAG
GAC TGT AGC AGT GAA AAG GCT GAA CAA CAG TGG GCT CTT TAT GCA GAT GGT
40 TCA ATA CGT CCT CAG CAA AAC CGA GAT AAT TGC CTT ACA AGT GAT TCT AAT
ATA CCG GAA ACA GTT GTT AAG ATC CTC TCT TGT GGC CCT GCA TCC TCT GGC
45 CAA CGA TGG ATG TTC AAG AAT GAT GGA ACC ATT TTA AAT TTG TAT AGT GGA
TTG GTG TTA GAT GTG AGG CGA TCG GAT CCG AGC CTT AAA CAA ATC ATT CTT
50 TAC CCT CTC CAT GGT GAC CCA AAC CAA ATA TGG TTA CCA TTA TTT
55

or a nucleotide sequence which is equivalent to said sequence or portion by virtue of the degeneracy of the genetic code.

60 3. A recombinant DNA molecule according to Claim 1 or Claim 2 characterised in that it includes at the NH₂- terminal end DNA coding for a signal polypeptide having the nucleotide sequence

65

EP 0 145 111 B1

ATG TAT GCA GTG GCA ACA TGG CTT TGT TTT GCA

5

TCC ACC TCA GGG TGG TCT TTC ACA TTA GAG

GAT AAC AAC

10 or a nucleotide sequence which is equivalent thereto by virtue of the degeneracy of the genetic code.
4. A recombinant DNA molecule according to Claim 1, characterised in that it includes as an insert the following DNA sequence which codes for preproridin

ATG TAT GCA GTG GCA ACA TGG CTT

15

TGT TTG GAA TCC ACC TCA GGG TGG TCT TTC ACA TTA GAG GAT AAC AAC ATA

20

TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGT GCC ACT GTG

CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TTA ACA ACT GGA

25

GCT GAT GTG AGA CAT GAT ATA CCA GTG TIG CCA AAC AGA GTT GGT TTG CCT

30

ATA AAC CAA CGG TTT ATT TTA GTT GAA CTC TCA AAT CAT GCA GAG CTT TCT

GTG ACA TTA GCC CTG GAT GTC ACC AAT GCA TAT GTG GTC GGC TAC CGT GCT

35

GGA AAT AGC GCA TAT TTC TTT CAT CCT GAC AAT CAG GAA GAT GCA GAA GCA

ATC ACT CAT CTT TTC ACT GAT GTT CAA AAT CGA TAT ACA TTC GCC TTT GGT

40

GGT AAT TAT GAT AGA CTT GAA CAA CTT GCT GGT AAT CTG AGA GAA AAT ATC

GAG TTG GGA AAT GGT CCA CTA GAG GAG GCT ATC TCA GCG CTT TAT TAT TAC

45

AGT ACT GGT GGC ACT CAG CTT CCA ACT CTG GCT CGT TCC TTT ATA ATT TGC

50

ATC CAA ATG ATT TCA GAA GCA GCA AGA TTC CAA TAT ATT GAG GGA GAA ATG

CGC ACG AGA ATT AGG TAC AAC CGG AGA TCT GCA CCA GAT CCT AGC GTA ATT

55

ACA CTT GAG AAT AGT TGG GGG AGA CTT TCC ACT GCA ATT CAA GAG TCT AAC

CAA GGA GCC TTT GCT AGT CCA ATT CAA CTG CAA AGA CGT AAT GGT TCC AAA

60

TTC AGT GTG TAC GAT GTG AGT ATA TTA ATC CCT ATC ATA GCT CTC ATG GTG

65

TAT AGA TGC GCA CCT CCA CCA TCG TCA CAG TTT TCT TTG CTT ATA AGG CCA

EP 0 145 111 B1

GTG GTA CCA AAT TTT AAT GCT GAT GTT TGT ATG GAT CCT GAG CCC ATA GTG
 5 CGT ATC GTA GGT CGA AAT GGT CTA TGT GTT GAT GTT AGG GAT GGA AGA TTC
 CAC AAC GGA AAC GCA ATA CAG TTG TGG CCA TGC AAG TCT AAT ACA GAT GCA
 10 AAT CAG CTC TGG ACT TTG AAA AGA GAC AAT ACT ATT CGA TCT AAT GGA AAG
 TGT TTA ACT ACT TAC GGG TAC AGT CCG GGA GTC TAT GTG ATG ATC TAT GAT
 15 TGC AAT ACT GCT GCA ACT GAT GCC ACC CGC TGG CAA ATA TGG GAT
 AAT GGA ACC ATC ATA AAT CCC AGA TCT AGT CTA GTT TTA GCA GCG ACA TCA
 20 GGG AAC AGT GGT ACC ACA CTT ACG GTG CAA ACC AAC ATT TAT GCC GTT AGT
 25 CAA GGT TGG CIT CCT ACT AAT AAT ACA CAA CCT TTT GTT ACA ACC ATT GTT
 GGG CTA TAT GGT CTG TGC TTG CAA GCA AAT AGT GGA CAA GTA TGG ATA GAG
 30 GAC TGT AGC AGT GAA AAG GCT GAA CAA CAG TGG GCT CTT TAT GCA GAT GGT
 TCA ATA CGT CCT CAG CAA AAC CGA GAT AAT TGC CTT ACA AGT GAT TCT AAT
 35 ATA CGG GAA ACA GTT GTT AAG ATC CTC TCT TGT GGC CCT GCA TCC TCT GGC
 40 CAA CGA TGG ATG TTC AAG AAT GAT GGA ACC ATT TTA AAT TTG TAT AGT GGA
 TTG GTG TTA GAT GTG AGG CGA TCG GAT CCG AGC CTT AAA CAA ATC ATT CTT
 45 TAC CCT CTC CAT GGT GAC CCA AAC CAA ATA TGG TTA CCA TTA TTT

or a nucleotide sequence which is equivalent to said sequence or portion by virtue of the degeneracy of the genetic code.

50 5. A recombinant DNA molecule according to any preceding claim characterised in that said insert has been introduced into a cloning vector which is a plasmid or a bacteriophage to obtain said recombinant DNA molecule.

55 6. A modified *E. Coli* or *Saccharomyces cerevisiae* host cell containing a recombinant DNA molecule according to any preceding claim.

7. A modified host microorganism according to Claim 6, when appendant to Claim 4, wherein the host is *E. Coli*, characterised in that the cloning vector is plasmid selected from pBR322, pAT153, pUC8, pGS15 or pMB9.

60 8. A modified host microorganism according to Claim 6, when appendant to Claim 4, wherein the host cell is *Saccharomyces Cerevisiae*, characterised in that the cloning vector is pMa91, pMA230, YRp7, pLC544 and YEp6.

9. A method of preparing cDNA having a nucleotide sequence as defined in any of Claims 1 to 4 characterised by separating from a mixture of mRNA's obtained from *Ricinus Communis* tissue a molecular weight fraction containing mRNA coding for preproinsulin and synthesising cDNA from this by reverse transcription.

EP 0 145 111 B1

10. A method of obtaining a recombinant DNA molecule by inserting a double-stranded cDNA according to Claim 8 into a cloning vector.

11. A method of obtaining a genetically-modified *E. Coli* or *Saccharomyces cerevisiae* host characterised in that a recombinant DNA molecule according to Claim 9 is introduced into said host.

Patentansprüche

1. Ein rekombinantes DNS-Molekül, dadurch gekennzeichnet, daß es als Einfügung DNS enthält, welche die A-Kette des Ricins kodiert und die folgende Nukleotid-Sequenz oder einen Teil davon besitzt, welche ein Polypeptid kodiert, das die Ribosomen-Inaktivitäts-Eigenschaft der A-Kette des Ricins zeigt:

ATA TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGI GCC ACT GTG

CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TIA ACA ACT GGA

GCT GAT GTG AGA CAT GAT ATA CCA GTG TIG CCA AAC AGA GTT GGT TTG CCT

ATA AAC CAA CGG TTT ATT TTA GTT GAA CTC TCA AAT CAT GCA GAG CTT TCT

GTT ACA TTA GCC CTG GAT GTC ACC AAT GCA TAT GTG GTC GGC TAC CGT GCT

GGA AAT AGC GCA TAT TTC TTT CAT CCT GAC AAT CAG GAA GAT GCA GAA GCA

ATC ACT CAT CTT TTC ACT GAT GTT CAA AAT CGA TAT ACA TTC GCC TTT GGT

GAG TTG GGA AAT GGT CCA CTA GAG GAG GCT ATC TCA GCG CTT TAT TAT TAC

AGT ACT GGT GGC ACT CAG CTT CCA ACT CTG GCT CGT TCC TTT ATA ATT TGC

ATC CAA ATG ATT TCA GAA GCA GCA AGA TTC CAA TAT ATT GAG GGA GAA ATG

CGC ACG AGA ATT AGG TAC AAC CGG AGA TCT GCA CCA GAT CCT AGC GTA ATT

ACA CTT GAG AAT AGT TGG GGG AGA CTT TCC ACT GCA ATT CAA GAG TCT AAC

CAA GGA GCC TTT GCT AGT CCA ATT CAA CTG CAA AGA CGT AAT GGT TCC AAA

TTT AGT GTG TAC GAT GTG AGT ATA TTA ATC CTT ATC ATA GCT CTC ATG GTG

TAT AGA TGC GCA CCT CCA CCA TCG TCA CAG TTT

oder eine Nukleotid-Sequenz, die wegen der Degenerierung des genetischen Codes dieser Sequenz oder diesem Teil äquivalent ist.

2. Rekombinantes DNS-Molekül nach Anspruch 1, dadurch gekennzeichnet, daß es als Einfügung DNS enthält, die Ricins kodiert und die folgende Nukleotid-Sequenz oder einen Teil davon besitzt, welche ein Protein kodiert, das die Zell-Bindungs- und Ribosomen-Inaktivitäts-Eigenschaft des Ricins zeigt:

ATA TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGI GCC ACT GTG

CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TIA ACA ACT GGA

EP 0 145 111 B1

GCT GÀT GTG AGA CAÏ GAT ATA CCA ÒTG TIG CCA AÀC AGA GTT GGT TTG CCT
ATA ÀAC CAA CGG TTT ATT TTA GTÏ GAA CTC TCA ÀAT CAT GCA GÀG CTT TCT
GTÏ ACA TTA GCC ÒTG GAT GTC ÀCC AAT GCA TAÏ GTG GTC GGC ÌAC CGT GCT
GGA AAT AGC GCÀ TAT TTC TTT ÇAT CCT GAC AÀT CAG GAA GAÏ GCA GAA GCA
ÀTC ACT CAT CÏT TTC ACT GAÏ GTT CAA AAT ÇGA TAT ACA TÏC GCC TTT GGT
GGT AAT TAT ÒAT AGA CTT GÀA CAA CTT GCT GGT AAT CTG ÀGA GAA AAT ÀTC
GAG TTG GGÀ AAT GGT CCA ÇTA GAG GAG GÇT ATC TCA GCÇ CTT TAT TAT ÌAC
AGT ACT GGT GGC ACT CAG CTT CCA ACT ÒTG GCT CGT TÇC TTT ATA ATT TGC
ATC CAA ÀTG ATT TCA GÀA GCA GCA AGÀ TTC CAA TAT ÀTT GAG GGA GÀA ATG
CGC ACÇ AGA ATT AGG ÌAC AAC CGG ÀGA TCT GCA CCÀ GAT CCT AGC ÒTA ATT
ACA CÏT GAG AAT AGÏ TGG GGG AGA ÇTT TCC ACT GÇA ATT CAA GAÇ TCT AAC
CAA ÒGA GCC TTT GÇT AGT CCA ATT CAA CTG CAA ÀGA CGT AAT GGT TCC AAA
TTÇ AGT GTG TAC ÒAT GTG AGT AÏA TTA ATC CCT ATC ATA GCT ÇTC ATG GTG
TÀT AGA TGC GCÀ CCT CCA CCA ÌCG TCA CAG TÏT TCT TTG CÏT ATA AGG CCA
ÒTG GTA CCA AÀT TTT AAT GCT GAT GTT TGT ÀTG GAT CCT GÀG CCC ATA GTÇ
CGT ATC GTA GGT CGA AAT GGT CTA TGT GTÏ GAT GTT AGG ÒAT GGA AGA TÏC
CAC AAC GGÀ AAC GCA ATA ÇAG TTG TGG CÇA TGC AAG TCT AAT ACA GAT ÒCA
AAT CAG CTC TGG ACT TTÇ AAA AGA GAC ÀAT ACT ATT CÇA TCT AAT GGÀ AAG
TGT TTA ÀCT ACT TAC GGG TAC AGT CCÇ GGA GTC TAT GTG ATG ATC TÀT GAT
TGC AAT ACT GCT GCA ÀCT GAT GCC ÀCC CGC TGG CAÀ ATA TGG GAT
ÀAT GGA ACC ÀTC ATA AAT CCÇ AGA TCT AGT ÇTA GTT TTA GÇA GCG ACA TÇA
GGG AAC AGT ÒGT ACC ACA CÏT ACG GTG CAÀ ACC AAC ATT ÌAT GCC GTT AGT

EP 0 145 111 B1

CAA GGT TGG CIT CCT ACT AAT AAT ACA CAA CCT TTT GTT ACA ACC ATT GTT
GGG CTA TAT GGT CTG TGC TTG CAA GCA AAT AGT GGA CAA GTA TGG ATÀ GAG
5 GAC TGT AGC AGT GAA AAG GCT GAA CAÀ CAG TGG GCT CTT TAT GCA GAT GGT
TCA ATÀ CGT CCT CAG CAA AAC CGA GAT AAT TGC CTT ACA AGT GAT TCT AAT
10 ATA CGG GAA ACA GTT GTT AAG ATC CTC TCT TGT GGC CCT GCA TCC TCT GGC
CAA CGA TGG ATG TTC AAG AAT GAT GGA ACC ATT TTA AAT TTG TAT AGT GGA
15 TTG GTG TTA GAT GTG AGG CGA TCG GAT CCG AGC CTT AAA CAA ATC ATT CTT
20 TAC CCT CTC CAT GGT GAC CCA AAC CAA ATA TGG TTA CCA TTA TTT

oder eine Nukleotid-Sequenz, die wegen der Degenerierung des genetischen Codes dieser Sequenz oder
25 diesem Teil äquivalent ist.

3. Rekombinantes DNS-Molekül nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß es am NH₂-
terminalen Ende DNS enthält, welche eine Signal-Polypeptid kodiert, welches die Nukleotid-Sequenz

30 ATG TAT GCA GTG GCA ACA TGG CTT TGT TTT GCA
TCC ACC TCA GGG TGG TCT TTC ACA TTA GAG
35 GAT AAC AAC

besitzt oder eine Nukleotid-Sequenz, die wegen der Degenerierung des genetischen Codes dazu äquivalent
ist.

40 4. Rekombinantes DNS-Molekül nach Anspruch 1, dadurch gekennzeichnet, daß es als Einfügung die
folgende DNS-Sequenz enthält, welche Preproridin kodiert:

45 ATG TAT GCA GTG GCA ACA TGG CTT
TGT TIT GGA TCC ACC TCA GGG TGG TCT TTC ACA TTA GAG GAT AAC AAC ATA
50 TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGI GCC ACT GTG
CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TTA ACA ACT GGA
55 GCT GAT GTG AGA CAT GAT ATA CCA GTG TIG CCA AAC AGA GTT GGT TTG CCT
ATA AAC CAA CGG TTT ATT TTA GTT GAA CTC TCA AAT CAT GCA GAG CTT TCT
60 GTT ACA TTA GCC CTG GAT GTC ACC AAT GCA TAT GTG GTC GGC TAC CGT GCT
GGA AAT AGC GCA TAT TTC TTT CAT CCT GAC AAT CAG GAA GAT GCA GAA GCA
65

EP 0 145 111 B1

ATC ACT CAT CTT TTC ACT GAT GTT CAA AAT CGA TAT ACA TTC GCC TTT GGT
GGT AAT TAT GAT AGA CTT GAA CAA CTT GCT GGT AAT CTG AGA GAA AAT ATC
GAG TTG GGA AAT GGT CCA CTA GAG GAG GCT ATC TCA GCG CTT TAT TAT TAC
AGT ACT GGT GGC ACT CAG CTT CCA ACT CTG GCT CGT TCC TTT ATA ATT TGC
ATC CAA ATG ATT TCA GAA GCA GCA AGA TTC CAA TAT ATT GAG GGA GAA ATG
CGC ACG AGA ATT AGG TAC AAC CGG AGA TCT GCA CCA GAT CCT AGC GTA ATT
ACA CTT GAG AAT AGT TGG GGG AGA CTT TCC ACT GCA ATT CAA GAG TCT AAC
CAA GGA GCC TTT GCT AGT CCA ATT CAA CTG CAA AGA CGT AAT GGT TCC AAA
TTC AGT GTG TAC GAT GTG AGT ATA TTA ATC CCT ATC ATA GCT CTC ATG GTG
TAT AGA TGC GCA CCT CCA CCA TCG TCA CAG TTT TCT TTG CTT ATA AGG CCA
GTG GTA CCA AAT TTT AAT GCT GAT GTT TGT ATG GAT CCT GAG CCC ATA GTG
CGT ATC GTA GGT CGA AAT GGT CTA TGT GTT GAT GTT AGG GAT GGA AGA TTC
CAC AAC GGA AAC GCA ATA CAG TTG TGG CCA TGC AAG TCT AAT ACA GAT GCA
AAT CAG CTC TGG ACT TTG AAA AGA GAC AAT ACT ATT CGA TCT AAT GGA AAG
TGT TTA ACT ACT TAC GGG TAC AGT CCG GGA GTC TAT GTG ATG ATC TAT GAT
TGC AAT ACT GCT GCA ACT GAT GCC ACC CGC TGG CAA ATA TGG GAT
AAT GGA ACC ATC ATA AAT CCC AGA TCT AGT CTA GTT TTA GCA GCG ACA TCA
GGG AAC AGT GGT ACC ACA CTT ACG GTG CAA ACC AAC ATT TAT GCC GTT AGT
CAA GGT TGG CIT CCT ACT AAT AAT ACA CAA CCT TTT GTT ACA ACC ATT GTT
GGG CTA TAT GGT CTG TGC TTG CAA GCA AAT AGT GGA CAA GTA TGG ATA GAG
GAC TGT AGC AGT GAA AAG GCT GAA CAA CAG TGG GCT CTT TAT GCA GAT GGT
TCA ATA CGT CCT CAG CAA AAC CGA GAT AAT TGC CTT ACA AGT GAT TCT AAT

EP 0 145 111 B1

ATA CGG GAA ACA GTT GTT AAG ATC CTC TCT TGT GGC CCT GCA TCC TCT GGC

5 CAA CGA TGG ATG TTC AAG AAT GAT GGA ACC ATT TTA AAT TTG TAT AGT GGA

TTG GTG TTA GAT GTG AGG CGA TCG GAT CCG AGC CTT AAA CAA ATC ATT CTT

10 TAC CCT CTC CAT GGT GAC CCA AAC CAA ATA TGG TTA CCA TTA TTT

oder eine Nukleotid-Sequenz, die wegen der Degenerierung des genetischen Codes dieser Sequenz oder diesem Teil äquivalent ist.

15 5. Rekombinantes DNS-Molekül nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß die Einfügung in einen Klonierungs-Zwischenträger, der ein Plasmid oder ein Bakteriophage ist, eingeführt wurde, um das rekombinante DNS-Molekül zu erhalten.

6. Modifizierte E. Coli oder Saccharomyces Cervisiae-Wirtszelle, welche ein rekombinantes DNS-Molekül nach einem der vorhergehenden Ansprüche enthält.

20 7. Ein modifizierter Wirts-Mikroorganismus nach Anspruch 6 und 4, wobei die Wirts-Zelle E. Coli ist, dadurch gekennzeichnet, daß der Klonierungs-Zwischenträger ein Plasmid, ausgewählt aus pBR322, pAT153, pUC8, pGS15 oder pMB9 ist.

8. Ein modifizierter Wirts-Mikroorganismus nach Anspruch 6 und 4, wobei die Wirts-Zelle Saccharomyces Cervisiae ist, dadurch gekennzeichnet, daß der Klonierungs-Zwischenträger pMA91, 25 pMA230, YRp7, pLC544 and YEp6 ist.

9. Methode zur Darstellung von cDNS mit einer Nukleotid-Sequenz wie in einem der Ansprüche 1 bis 4 definiert, gekennzeichnet durch Abtrennen einer Molekular-gewichtsfraction, die Preprorcin kodierende mRNA enthält, aus einer Mischung von aus Ricinus Communis-Gewebe erhaltenen mRNA und durch 30 Synthetisieren von cDNS aus dieser mittels reverser Transkription.

10. Method, ein rekombinantes DNS-Molekül durch Einsetzen einer doppelsträngigen cDNS nach Anspruch 8 in einen Klonierungs-Zwischenträger zu erhalten.

11. Methode zur Erhaltung eines genetisch modifizierten E. Coli- oder Saccharomyces Cervisiae-Wirts, dadurch gekennzeichnet, daß ein rekombinantes DNS-Molekül nach Anspruch 9 in diesen Wirt eingeführt 35 wird.

35 Revendications

1. Molécule d'ADN recombinant caractérisée en ce qu'elle comprend en tant que segment d'insertion 40 un ADN codant pour la chaîne A de la ricine et ayant la séquence nucléotidique ci-dessus ou une partie de celle-ci codant pour un polypeptide qui manifeste la propriété d'inactivation ribosomale de la chaîne A de la ricine:

ATA

45 TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGI GCC ACT GTG

CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TIA ACA ACT GGA

50 GCT GAT GTG AGA CAT GAT ATA CCA GTG TIG CCA AAC AGA GTT GGT TTG CCT

55 ATA AAC CAA CGG TTT ATT TTA GTT GAA CTC TCA AAT CAT GCA GAG CTT TCT

GGT ACA TTA GCC CTG GAT GTC ACC AAT GCA TAT GTG GTC GGC TAC CGT GCT

60 GGA AAT AGC GCA TAT TTC TTT CAT CCT GAC AAT CAG GAA GAT GCA GAA GCA

65 ATC ACT CAT CTT TTC ACT GAT GTT CAA AAT CGA TAT ACA TTC GCC TTT GGT

EP 0 145 111 B1

GGT AAT TAT GAT AGA CTT GAA CAA CTT GCT GGT AAT CTG AGA GAA AAT ATC
5 GAG TTG GGA AAT GGT CCA CTA GAG GAG GCT ATC TCA GCG CTT TAT TAT TAC
AGT ACT GGT GGC ACT CAG CTT CCA ACT CTG GCT CGT TCC TTT ATA ATT TGC
10 ATC CAA ATG ATT TCA GAA GCA GCA AGA TTC CAA TAT ATT GAG GGA GAA ATG
CGC ACG AGA ATT AGG TAC AAC CGG AGA TCT GCA CCA GAT CCT AGC GTA ATT
15 ACA CTT GAG AAT AGT TGG GGG AGA CTT TCC ACT GCA ATT CAA GAG TCT AAC
20 CAA GGA GCC TTT GCT AGT CCA ATT CAA CTG CAA AGA CGT AAT GGT TCC AAA
TTC AGT GTG TAC GAT GTG AGT ATA TTA ATC CCT ATC ATA GCT CTC ATG GTG
25 TAT AGA TGC GCA CCT CCA CCA TCG TCA CAG TTT

ou une séquence nucléotidique qui est équivalente à ladite séquence ou partie de la séquence en raison de la dégénérescence du code génétique.

30 2. Molécule d'ADN recombinant selon la revendication 1, caractérisée en ce qu'elle comprend en tant que segment d'insertion un ADN codant pour la ricine et ayant la séquence nucléotidique ci-dessous ou une portion de celle-ci codant pour un protéine qui manifeste la propriété de liaison aux cellules et d'inactivation ribosomale de la ricine:

35 ATA
TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGI GCC ACT GTG
40 CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TTA ACA ACT GGA
GCT GAT GTG AGA CAT GAT ATA CCA GTG TIG CCA AAC AGA GTT GGT TTG CCT
45 ATA AAC CAA CGG TTT ATT TTA GTT GAA CTC TCA AAT CAT GCA GAG CTT TCT
50 GTT ACA TTA GCC CTG GAT GTC ACC AAT GCA TAT GTG GTC GGC TAC CGT GCT
GGA AAT AGC GCA TAT TTC TTT CAT CCT GAC AAT CAG GAA GAT GCA GAA GCA
55 ATC ACT CAT CTT TTC ACT GAT GTT CAA AAT CGA TAT ACA TTC GCC TTT GGT
GGT AAT TAT GAT AGA CTT GAA CAA CTT GCT GGT AAT CTG AGA GAA AAT ATC
60 GAG TTG GGA AAT GGT CCA CTA GAG GAG GCT ATC TCA GCG CTT TAT TAT TAC
AGT ACT GGT GGC ACT CAG CTT CCA ACT CTG GCT CGT TCC TTT ATA ATT TGC
65

EP 0 145 111 B1

ATC CAA ATG ATT TCA GAA GCA GCA AGA TTC CAA TAT ATT GAG GGA GAA ATG
CGC ACG AGA ATT AGG TAC AAC CGG AGA TCT GCA CCA GAT CCT AGC GTA ATT
ACA CTT GAG AAT AGT TGG GGG AGA CTT TCC ACT GCA ATT CAA GAG TCT AAC
CAA GGA GCC TTT GCT AGT CCA ATT CAA CTG CAA AGA CGT AAT GGT TCC AAA
TTC AGT GTG TAC GAT GTG AGT ATA TTA ATC CCT ATC ATA GCT CTC ATG GTG
TAT AGA TGC GCA CCT CCA CCA TCG TCA CAG TTT TCT TTG CTT ATA AGG CCA
GTG GTA CCA AAT TTT AAT GCT GAT GTT TGT ATG GAT CCT GAG CCC ATA GTG
CGT ATC GTA GGT CGA AAT GGT CTA TGT GTT GAT GTT AGG GAT GGA AGA TTC
CAC AAC GGA AAC GCA ATA CAG TTG TGG CCA TGC AAG TCT AAT ACA GAT GCA
AAT CAG CTC TGG ACT TTG AAA AGA GAC AAT ACT ATT CGA TCT AAT GGA AAG
TGT TTA ACT ACT TAC GGG TAC AGT CCG GGA GTC TAT GTG ATG ATC TAT GAT
TGC AAT ACT GCT GCA ACT GAT GCC ACC CGC TGG CAA ATA TGG GAT
AAT GGA ACC ATC ATA AAT CCC AGA TCT AGT CTA GTT TTA GCA GCG ACA TCA
GGG AAC AGT GGT ACC ACA CTT ACG GTG CAA ACC AAC ATT TAT GCC GTT AGT
CAA GGT TGG CIT CCT ACT AAT AAT ACA CAA CCT TTT GTT ACA ACC ATT GTT
GGG CTA TAT GGT CTG TGC TTG CAA GCA AAT AGT GGA CAA GTA TGG ATA GAG
GAC TGT AGC AGT GAA AAG GCT GAA CAA CAG TGG GCT CTT TAT GCA GAT GGT
TCA ATA CGT CCT CAG CAA AAC CGA GAT AAT TGC CTT ACA AGT GAT TCT AAT
ATA CCG GAA ACA GTT GTT AAG ATC CTC TCT TGT GGC CCT GCA TCC TCT GGC
CAA CGA TGG ATG TTC AAG AAT GAT GGA ACC ATT TTA AAT TTG TAT AGT GGA
TTG GTG TTA GAT GTG AGG CGA TCG GAT CCG AGC CTT AAA CAA ATC ATT CTT
TAC CCT CTC CAT GGT GAC CCA AAC CAA ATA TGG TTA CCA TTA TTT

EP 0 145 111 B1

ou une séquence nucléotidique qui est équivalente à ladite séquence ou partie de la séquence ou partie de celle-ci en raison de la dégénérescence du code génétique.

3. Molécule d'ADN recombinant selon la revendication 1 ou la revendication 2, caractérisée en ce qu'elle comprend à l'extrémité NH₂-terminale un ADN codant pour un polypeptide de signal ayant la séquence nucléotidique:

ATG TAT GCA GTG GCA ACA TGG CTT TGT TTT GCA

10

TCC ACC TCA GGG TGG TCT TTC ACA TTA GAG

GAT AAC AAC

15

ou une séquence nucléotidique qui est équivalente à celle-ci en raison de la dégénérescence du code génétique.

4. Molécule d'ADN recombinant selon la revendication 1, caractérisée en ce qu'elle comprend en tant que segment d'insertion la séquence d'ADN suivante qui code pour la préproinsuline:

20

ATG TAT GCA GTG GCA ACA TGG CTT

TGT TTT GGA TCC ACC TCA GGG TGG TCT TTC ACA TTA GAG GAT AAC AAC ATA

25

TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA GCG GGT GCC ACT GTG

30

CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC GGT CGT TTA ACA ACT GGA

GCT GAT GTG AGA CAT GAT ATA CCA GTG TIG CCA AAC AGA GTT GGT TTG CCT

35

ATA AAC CAA CGG TTT ATT TTA GTT GAA CTC TCA AAT CAT GCA GAG CTT TCT

GTT ACA TTA GCC CTG GAT GTC ACC AAT GCA TAT GTG GTC GGC TAC CGT GCT

40

GGA AAT AGC GCA TAT TTC TTT CAT CCT GAC AAT CAG GAA GAT GCA GAA GCA

45

ATC ACT CAT CTT TTC ACT GAT GTT CAA AAT CGA TAT ACA TTC GCC TTT GGT

GGT AAT TAT GAT AGA CTT GAA CAA CTT GCT GGT AAT CTG AGA GAA AAT ATC

50

GAG TTG GGA AAT GGT CCA CTA GAG GAG GCT ATC TCA GCG CTT TAT TAT TAC

AGT ACT GGT GGC ACT CAG CTT CCA ACT CTG GCT CGT TCC TTT ATA ATT TGC

55

ATC CAA ATG ATT TCA GAA GCA GCA AGA TTC CAA TAT ATT GAG GGA GAA ATG

CGC ACG AGA ATT AGG TAC AAC CGG AGA TCT GCA CCA GAT CCT AGC GTA ATT

60

ACA CTT GAG AAT AGT TGG GGG AGA CTT TCC ACT GCA ATT CAA GAG TCT AAC

65

CAA GGA GCC TTT GCT AGT CCA ATT CAA CTG CAA AGA CGT AAT GGT TCC AAA

EP 0 145 111 B1

TTC AGT GTG TAC GAT GTG AGT ATA TTA ATC CCT ATC ATA GCT CTC ATG GTG
 5 TAT AGA TGC GCA CCT CCA CCA TCG TCA CAG TTT TCT TTG CTT ATA AGG CCA
 GTG GTA CCA AAT TTT AAT GCT GAT GTT TGT ATG GAT CCT GAG CCC ATA GTG
 10 CGT ATC GTA GGT CGA AAT GGT CTA TGT GTT GAT GTT AGG GAT GGA AGA TTC
 CAC AAC GGA AAC GCA ATA CAG TTG TGG CCA TGC AAG TCT AAT ACA GAT GCA
 15 AAT CAG CTC TGG ACT TTG AAA AGA GAC AAT ACT ATT CGA TCT AAT GGA AAG
 TGT TTA ACT ACT TAC GGG TAC AGT CCG GGA GTC TAT GTG ATG ATC TAT GAT
 20 TGC AAT ACT GCT GCA ACT GAT GCC ACC CGC TGG CAA ATA TGG GAT
 25 AAT GGA ACC ATC ATA AAT CCC AGA TCT AGT CTA GTT TTA GCA GCG ACA TCA
 GGG AAC AGT GGT ACC ACA CTT ACG GTG CAA ACC AAC ATT TAT GCC GTT AGT
 30 CAA GGT TGG CIT CCT ACT AAT AAT ACA CAA CCT TTT GTT ACA ACC ATT GTT
 GGG CTA TAT GGT CTG TGC TTG CAA GCA AAT AGT GGA CAA GTA TGG ATA GAG
 35 GAC TGT AGC AGT GAA AAG GCT GAA CAA CAG TGG GCT CTT TAT GCA GAT GGT
 40 TCA ATA CGT CCT CAG CAA AAC CGA GAT AAT TGC CTT ACA AGT GAT TCT AAT
 ATA CGG GAA ACA GTT GTT AAG ATC CTC TCT TGT GGC CCT GCA TCC TCT GGC
 45 CAA CGA TGG ATG TTC AAG AAT GAT GGA ACC ATT TTA AAT TTG TAT AGT GGA
 50 TTG GTG TTA GAT GTG AGG CGA TCG GAT CCG AGC CTT AAA CAA ATC ATT CTT
 TAC CCT CTC CAT GGT GAC CCA AAC CAA ATA TGG TTA CCA TTA TTT

55 ou une séquence nucléotidique qui est équivalente à ladite séquence ou partie de celle-ci en raison de la dégénérescence du code génétique.

5. Molécule d'ADN recombinant selon l'une quelconque des revendications précédentes, caractérisée en ce que ledit segment d'insertion a été introduit dans un vecteur de clonage qui est un plasmide ou un bactériophage pour obtenir ladite molécule d'ADN recombinant.

60 6. Cellule hôte modifiée d'*E. coli* ou de *Saccharomyces Cervisiae* contenant une molécule d'ADN recombinant selon l'une quelconque des revendications précédentes.

7. Microorganisme hôte modifié selon la revendication 6, lorsqu'elle est dépendante de la revendication 4, dans laquelle la cellule hôte est *E. coli*, caractérisé en ce que le vecteur de clonage est un plasmide choisi parmi pBR322, pAT153, pUC8, pGS15 ou pMB.

65 8. Microorganisme hôte modifié selon la revendication 6, lorsqu'elle est dépendante de la

EP 0 145 111 B1

revendication 4, dans laquelle la cellule hôte est *Sacchoromyces Cervisiae*, caractérisé en ce que le vecteur d clonage est pMa91, pMA230, YRp7, pLC544 et YEp6.

9. Procédé de préparation d'ADNc ayant une séquence nucléotidique telle que définie dans l'une quelconque des revendications 1 à 4, caractérisé en ce que l'on sépare à partir d'un mélange d'ARNm obtenu à partir de tissu de *Ricinus Communis* une fraction de poids moléculaire contenant l'ARNm codant pour la préproricine et en ce que l'on synthétise l'ADNc à partir de celui-ci par transcription inverse.

10. Procédé d'obtention d'une molécule d'ADN recombinant par insertion d'un ADNc double brin selon la revendication 8 dans un vecteur de clonage.

11. Procédé d'obtention d'un hôte génétiquement modifié d'*E. coli* ou *Sacchoromyces cervisiae*, caractérisé en ce que l'on introduit une molécule d'ADN recombinant selon la revendication 9 dans ledit hôte.

15

20

25

30

35

40

45

50

55

60

65

